2015-08-05

Development of multi-probe fluorescence-based assay for boNTA detection

Dadgar, Saedeh

Development of Multi-Probe Fluorescence-Based Assay for BoNTA Detection

A thesis presented to

The Faculty of Graduate Studies

of

Lakehead University

by

SAEDEH DADGAR

In partial fulfillment of the requirements
for the degree of
Doctor of Philosophy in Chemistry and Materials Science

December 18, 2013

Abstract

Botulinum neurotoxins (BoNTs) cause the lethal disease botulism through the inhibition of acetyl choline secretion by the cleavage of crucial SNARE proteins. Determination of critical residues in the protein sequence of BoNT serotype A was the primary step to identify novel fluorescence recognition agents for BoNTA. Computational and experimental studies were employed to identify paclitaxel as a new inhibitor (IC_{50} equal to 5.2 μM) for the proteolytic activity of BoNTA light chain (LC) using Fluorescence Resonance Energy Transfer (FRET) assay. A fluorescent derivative of paclitaxel (PAC-BDP) exhibited binding to complex BoNTA. A Primary Amines Database comprised of 1,153 compounds suitable for fluorescent labeling was computationally screened to select 6-aminofluorescein (6-AFLU) and aspartame (APM) as recognition agent candidates. Fluorescent labeled APM (APM-BDP) was synthesised and the purity of the compound was confirmed using liquid chromatography mass spectrometry (LC/MS) and nuclear magnetic resonance (NMR). 6-AFLU exhibited good binding affinity to BoNTA heavy chain (HC) with an EC_{50} of 546 ± 60nM, whereas APM-BDP displayed binding to BoNTA LC with an EC_{50} of 20.96 ± 10 nM, as determined by fluorescence polarization (FP) assay. APM was shown to compete with APM-BDP for the same binding site in BoNTA LC, but showed no binding to BoNTA HC in FP competition assay. Also, aminopterin (AMN) and 6-AFLU exhibited binding to the same site of BoNTA HC, whereas desmosine (DES) showed affinity to a different binding site in BoNTA HC. Additionally, PAC exhibited binding to BoNTA LC, however paclitaxel (PAC) did not compete with APM-BDP for the same binding region. PAC-BDP showed binding to both

BoNTA LC and BoNTA HC and did not compete with APM-BDP for the same binding site in BoNTA LC. A library consisting of 1,624 commercially available radiolabeled ligands were screened computationally to select the ligands with binding affinity against BoNTA LC and HC. The binding of [^{3}H] Aminopterin and [^{3}H] desmosine was shown to be concentration-dependent with $EC_{50}$of 703 ± 98 nM and 1.6 ± 0.3μM, respectively, against BoNTA HC using scintillation proximity assay (SPA). [^{3}H]Solanesyl pyrophosphate (Solanesyl PP) exhibited high binding to both BoNTA LC and BoNTA HC. However its related compound, [^{3}H]Solanesol, show no binding against BoNTA LC or BoNTA HC using SPA assay.

The development of a fast, simple, reliable assay for BoNTA detection is essential since mouse lethality assay (MLA), the only trustable assay, is a costly, time consuming and complicated assay. In addition, detection of BoNTA in the initial steps of contamination is critical for successful treatment. This study demonstrated that FP can be used as a platform for BoNTA detection and that PAC-BDP, APM-BDP and 6-AFLU can be used simultaneously since they bind to different binding regions of BoNTA. The identified recognition agents can potentially be used in a multi-probe FP assay against the whole BoNTA complex.

Acknowledgments

I am deeply grateful to my supervisor Dr. Wely B. Floriano for her guidance, encouragement, continuous support, and allowing me to conduct such an exciting project in my Ph.D. study. I would also like to extend my sincere thanks to the members of my supervisory committee Dr. Michael Campbell and Dr. Robert C. Mawhinney, and the external examiner, Dr. Peter Oelschlaeger, for sharing their valuable time by lending their knowledge and support to my thesis. Additionally, I would like to thank Dr. Gregory Spivak coordinator of the Ph.D. program in Chemistry and Materials Science. I am also very grateful to Dr. Christopher P. Phenix and Dr. Morshed Chowdhury for their great help on the synthesis of APM-BDP.

Finally, it is with much gratitude and sincerity I thank my husband and my two lovely kids, Niki and Daniel for their patience and encouragement during my Ph.D.

Dedication

To my wonderful husband, who has supported me in all my endeavors.

To the two lovely projects of my life: my daughter Niki and my son Daniel.

To my parents for their endless love and encouragement.

Table of Contents

LIST OF FIGURES:

LIST OF TABLES:

1 CHAPTER ONE: INTRODUCTION

1.1 INTRODUCTION

There are four important pathogen species in *Clostridium* genus including *Clostridium perfringens*, *Clostridium tetani, Clostridium difficile and Clostridium botulinum. C. botulinum* is an obligate anaerobic, gram positive and spore forming bacillus. These bacteria are found in soil and grow at 37 °C or 30 °C, pH 6.5-7.0. The bacterium [1] generates spores in unfavourable conditions to survive. *C. botulinum* produces toxin known as botulinum neurotoxin (BoNT). There are seven types of the toxin classified by the letters A through G. Human botulism is caused by toxin type A, B and E, whereas type F is implicated less often. Type C and D cause botulism in non-human species and type G never has been found to cause botulism in humans.[2]

1.1.1 Chemical structure of the toxin

BoNT is one of the most poisonous known toxins. Based on primate studies the lethal oral dose of BoNTA for a 70 kg human is estimated at 70 μg.[3] BoNT is a complex consisting of botulinum neurotoxin and nontoxic associated proteins (NAPs). The NAPs consist of hemagglutinins (HA) and a nontoxic nonhemagglutinin (NTNH). The number of NAPs is variable between two to six in different serotypes of BoNT. NAPs are protecting BoNTs from protease and acidic environment and also assist BoNT to cross the intestinal epithelial barrier. [4]All clostridial neurotoxins are synthesized by the bacteria as a single inactive polypeptide chain with a molecular weight about 150 kDa. These neurotoxins remain inactive after synthesis. Ultimately, they will be activated by bacterial or tissue protease which cleaves the BoNT to light

chain (LC) and heavy chain (HC). However, the light and heavy chains are still joined by a disulphide bond [1].

1.1.2 Mechanism of action of the toxin

The routes of administration of BoNT include ingestion, inhalation and wound. The BoNT starts its journey after entering to the body and absorbs through mucosal surfaces such as digestive and respiratory systems. Consequently, the BoNT enters the lymphatic system and blood circulation. Ultimately, BoNT reaches a nerve cell and binds to the ganglioside molecule on the cell surface of the nerve cells, which is followed by receptor-mediated endocytosis. The BoNT HC plays two roles including receptor binding through its C-terminus to nerve cells, and translocation of the light chain into the cytoplasm across the membrane through its N-terminus[5]. The BoNT LC exhibits zinc-endopeptidase activity and cleaves soluble NSF-attachment protein receptor (SNARE) proteins. SNARE proteins include synaptobrevin, syntaxin and SNAP-25. These proteins are essential for the secretion of neurotransmitters [6]. The synaptobrevins are located on synaptic vesicles and syntaxin and SNAP-25 are the plasma membrane proteins. Any SNARE proteins degradations disrupts formation of functional SNARE proteins complex and blocks acetylcholine's exocytosis and lead to paralysis. The target protein of BoNTA, C and E is SNAP-25 (synaptosomal associated protein) with 25 kDa molecular weight while the BoNTB, D, F and G cleave synaptobrevin as a target [2, 7].

1.1.3 Clinical symptoms

Double vision, dry mouth, dilated pupils, difficulty of speaking, muscle weakness, and flaccid paralysis are some early symptoms of botulism. Limb and respiratory muscle paralysis are the secondary symptoms of botulism, which can lead to death. The symptoms usually develop within 12 to 36 hours after BoNT ingestion or inhalation. Based on the amount of the

BoNT, recovery time will be different [8]. The current treatment of the BoNT is an antitoxin combined with respiratory supportive care. However, the treatment must be applied before BoNT reaches nerve cells because BoNT is not accessible after entering the nerve cell cytoplasm [9].

1.1.4 Clinical forms

Foodborne, wound, infant, adult infectious, and inadvertent are five clinical forms of botulism. Foodborne botulism is caused by ingestion of contaminated food by preformed toxin. In infant and adult infectious botulisms, ingested organisms produce toxin in the digestive system. The infant intestine does not have protective bacterial flora before the age of 6 months. The presence of clostridial spores in honey products has been reported; thus honey should not be given to children during the first year of life [10]. However, in wound botulism the organisms grow in deep wound and produce toxin. Inadvertent botulism occurs in people with movement disorders who were treated with intramuscular injections of BoNT [1-2].

1.1.5 BoNT detection methods

Many studies have been done to improve methods for BoNT detection due to the necessity of developing a rapid method to diagnose botulism. Table 1.1 summarizes some of the detection methods including mouse lethality assay (MLA), lateral flow assay (strip assay or Immuno-chromatographic assay), enzyme-linked immunoassays (ELISA), polymerase chain reaction (PCR), immuno-PCR detection, fluorescence polarization immunoassay (FPIA), cellular based assay and fluorescence polarization (FP) assay.

Table 1.1 Summary of some of the methods for detection of BoNT

Technique	Time	Sensitivity	Specificity	Cost	Suitable for HTS	Accuracy	Difficulties	Advantage or disadvantage	Ref.
Mouse Lethality Assay (MLA)	Slow (2-4 days)	The most sensitive assay	Identify all the seven serotypes	Expensive	No	Only accepted method	Laborious	Ethical issues	2, 8
Lateral Flow Assay	Fast (15 Min)	Low	Identify toxin types A,B, and E	Low cost		False positive (Due to food contamination)	Easy (Minimum training and equipment)	The results are qualitative (Yes/No)	11
Cellular Based Assay	Slow (2-3 days)		Able to detect active toxin	Expensive			Complex assay	Difficult to get quantitative results	2
ELISA	Fast (about 6 hours)	Less than MLA	Less than MLA	Expensive		False positive & False negative	Easy	High quality antibodies are not generally available	8
Polymerase Chain Reaction (PCR)	Fast	High	High	Expensive		-		PCR only detect BoNT gene, not active toxin	8
Immuno-PCR	Fast (Not as fast as ELISA)	Very sensitive (can reach to MLA)	High	Expensive		High concentration Of DNA-antibody Make non-specific PCR signal	More complex than ELISA	Useful to detect low level antigen	12
FPIA	Fast (Minimize the analysis Time)	Depends on binding affinity of the antibody		Expensive	Yes		Simple	Homogenous techniques	13
FP	Fast	High	High	Relatively inexpensive	Yes	Yes	Simple	Homogenous techniques & Adoptable to field use & Suitable for high throughput screening	14

1.1.6 Mouse lethality assay (MLA)

Mouse lethality assay is based on an intraperitoneal (IP) injection of a sample into laboratory mice. If the sample contains the toxin, mice show botulism symptoms. The minimum amount of the toxin which kills all the incubated mice is called mouse lethality dose (MLD). The sensitivity of the assay is 1 MLD for BoNTA, which is equal to 10 pg/ml of the toxin. The assay is the most sensitive method for BoNT detection, and also is able to identify all seven toxin serotypes. MLA has been used to detect the toxin indifferent samples such as serum, wound, gastric and food [8]. Although MLA is the only trustable assay for toxin detection, it has several disadvantages such as being costly, laborious, and slow as well as ethical issues (Table 1.1) [2].

1.1.7 Fluorescence polarization assay

FP is a powerful method that has been used for molecular interaction studies. This assay is based on movement and rotation of a fluorescent molecule in solution while the viscosity and temperature of the solution are constant. The fluorescence polarization is enhanced by increasing the solvent viscosity and decreasing the temperature. The emitted light of a large fluorescent molecule after excitation with plane polarized light, remains highly polarized due to slow movement of a large molecule in the excitation state. However, in the case of small fluorescent molecules the emitted light will be depolarized due to fast rotation of the molecule in the excitation state[15]. Therefore, changes in polarization intensity can be measured. The excitation state is the period of time between absorption of excitation photon and emission of a photon by the fluorescent molecule [14b, c]. The emitted polarization value for the free small fluorescent molecule will be low while the emitted polarization value for the small fluorescent molecule bound to a large molecule will be high [14c]. Fluorescence polarization can be calculated by the following equation:

mP = 1000 [FLUpar – FLUper] / [FLUpar + FLUper]

In this equation, P is polarization (given in unites of milli polarization) , FLUpar is the parallel fluorescence intensity measurement and FLUper is the perpendicular fluorescence intensity measurement. P (polarization) is independent of the fluorophore concentration and the value of emitted light.

FP is a homogenous assay that shows molecular interactions especially binding event in solution. Separation and washing steps are not required for FP assay making it faster, easier, cheaper and more accurate than other assay platforms for ligand binding. The reaction's equilibrium is disrupted in assays that require separation steps thus decreasing the assay accuracy [14c]. Also FP assay has high sensitivity and is suitable for high throughput screening. Fluorescence polarization measurements do not affect samples, so the samples can be analyzed and reanalyzed [14a, 14c, 16].

1.1.8 Fluorescence resonance energy transfer (FRET)

Another fluorescent based assay which has been used in many fields for molecular interaction studies is FRET assay. This assay utilizes fluorescent substrate to explore the binding interactions of molecules [7, 17]. Lately the FRET substrate peptide SNAPtide® was employed by several researchers to identify new inhibitors for BoNTA. SNAPtide is a synthetic peptide substrate and consists of a quencher and a fluorophore at the C-terminus and N-terminus, respectively. The cleavage site of SNAPtide is similar to SNAP-25 which is the native substrate of BoNTA. BoNTA LC cleaves SNAPtide and releases the fluorophore [7, 18]. The value of fluorescence intensity shows the enzymatic activity of BoNTA. In the presence of inhibitors, the inhibitor molecules bind to the active site of BoNTA LC thus decreasing the number of

unoccupied active sites available for substrate binding. Consequently, the fluorescence intensity is reduced by increasing the concentration of inhibitor [19].

1.1.9 Scintillation proximity assay

Scintillation proximity assay (SPA) is one of the homogenous, high sensitive and accurate methods to investigate molecular interactions in the biochemistry field, especially in enzyme and protein ligand interaction assays. This homogeneous assay only requires adding and mixing reagents as well as measuring the response without any separation or filtering steps[20]. However SPA is an expensive assay due to use of radio-isotopes. The principle of SPA technology is based on the distance between two molecules in solution. Prior to design SPA assay one of the involved molecules (protein) in the interaction is required to be labeled by biotin. The biotinylated molecule shows strong affinity to scintillant beads coated with streptavidin[20]. The interaction of radiolabeled molecule and biotinylated protein decreases the distance of scintillant beads and radiolabeled molecule. In this stage, decay of radiolabeled molecule can stimulate the scintillant contained inside the bead to emit light, and the emitted light can be measured by a scintillation counter. Either the ligand or the receptor is required to be radioactive labeled by one of the ^{125}I, ^{3}H, ^{14}C or ^{32}P radio-isotopes. The radioactive molecule decays particles and the particles travel different distance according to their environment and their energy[21].

1.1.10 Computer-aided molecular design

1.1.10.1 Virtual screening

Virtual screening has become an essential component of biochemistry and drug discovery. A variety of computational techniques are available to identify new active compounds

from chemical databases against any particular target. Ligand-based and structure-based are the two major approaches of virtual screening[22].

1.1.10.2 Pharmacophore modeling

A pharmacophore model is a 3D arrangement of essential functional groups for interaction with the target. The pharmacophore approaches facilitated identification of lead compounds in biochemistry and drug discovery [23]. A pharmacophore can be derived by employing ligand-based or structure-based pharmacophore modeling. The ligand-based pharmacophore modeling, structure-activity relationships (SAR) and quantitative structure-activity relationships (QSAR) are the main methods of ligand-based virtual screening approach. These methods are utilized to search for similarity in the structure of ligands to optimize the lead compounds. Ligands with similar structure to an active compound are often active themselves. The structure-based approaches depend on the knowledge of the 3D structure of the target protein free or bound to the ligand[24]. This valuable information is used for the molecular docking of ligands. Both ligand-based and structure-based pharmacophore models have been used broadly in virtual screening in recent years [23].

1.1.10.3 Molecular docking

Molecular docking enables us to predict binding interactions of protein-ligand, protein-DNA and protein-protein. This approach provides the ability of screening a large library of compounds for their binding against the target protein [25]. Molecular docking approach evaluates the interaction of potential ligands to the target and calculates the binding energy of ligand using molecular mechanics. Molecular docking searches for best conformation and orientation of ligand in the particular binding site and calculates binding affinity (scoring)[26,27].

*Hier*VLS is a computational approach to screen large libraries of chemical compounds rapidly. This approach enables us to determine the binding interaction of each ligand against each one of the available binding pockets of a protein. In this approach, the binding energy of each ligand to each binding region calculated and the ligands were ranked based on their binding energies to select best candidates. *Hier*VLS approach is comprised of the following steps including Level 0, Filter, Level 1 and Level 2. The numbers of configuration is generated for each ligand against each binding site of protein in Level 0. During the Filter step, the percentage of buried surface area and the number of hydrogen bonds between ligand and target calculated for each configuration. The configurations were ranked based on docking scores, buried surface area and the number of hydrogen bonds. The best numbers of configuration of each ligand were energy minimized at Level 1 and the configuration with lowest energy is selected for an all-atoms energy minimization at Level 2.

The Cassandra GUI [28] was used to import the input files (ligands and protein files) and organize the input and output files. *Hier*VLS comprise of the following major software programs, PASS, Dock 4.0, DMS, HBplus, Connolly, and Open Babel [29]. Open Babel is utilized to convert the file formats between programs. PASS program is used to identify potential binding pockets within the protein. Dock 4.0 is used to generate a large number of configurations (> 10,000) for each ligand bound to each binding region. The molecular surface, accessible surface area and number of hydrogen bonds of between protein and ligand are determined using DMS, Connolly, and HBplus software programs, respectively.

This method increases the accuracy of screening because the number of generated ligand-protein configurations for each ligand-protein pair decrease while the time required for computational calculation increases as the calculation progresses through the multiple steps of

HierVLS . The binding energy of the best configuration of each ligand calculated at the end. The force field binding energy for each binding region (*i)* calculated using the following equation:

$$FF\_BE_i = (PE + SE)_{BoundComplex} - (PE + SE)_{UnboundProtein} - (PE + SE)_{UnboundLigand}$$

Where FF_BE_i is force-field binding energy, PE represents potential energy, and SE is solvation energy. Compounds with more negative binding energy consider as potential active ligands. Additionally, the force-field binding energies were used to calculate Holistic Binding score using the following equation:

$$HBS_i = FF\_BE_i \times \frac{FF\_BE_i}{\sum_{j=1}^{n} FF\_BE_j}$$

Where HBS_i is the holistic binding score for a particular ligand bound to region i, and n represents the number of binding regions. The result of this approach is a set of ligands ranked by binding affinity for a particular binding site relative to all other potential sites within the target protein.

1.1.10.4Data processing and statistical analysis

In this thesis, all the data was analysed using Graphpad prism 6 (www.graphpad.com). Whenever appropriate, the experimental data was analysed using two way ANOVA. The two way ANOVA analysis is based on the assumption that the replicates are sampled from Gaussian distributions and determines how a response is influenced by two factors.

To determine whether the difference between two samples is statistically significant or not, we compared the mean of each sample (averaged over duplicate or triplicate measurements) to the mean of its related control in each row. There are two values that determine a test of significance, alpha and P value. Both alpha level and P value numbers are between zero and one.

Alpha level is the probability of getting the wrong result, and shows the level of confidence of the experiment. If the alpha level is set at 0.05, we would expect to get the same result 95 times out of 100 times repeating the experiment. The P value is the probability of getting the observed result. The smaller the p value, the greater is the statistical significance of the result. To conclude that the observed value did not occur by chance, the P value should be equal or less than the alpha level. In this study the alpha level was set at 0.05 to have the confidence interval of 95%. P value for each comparison was displayed as stars in the bar graphs. The four, three, two, and one stars represented P value lower than 0.0001, 0.001, 0.01 and 0.05, respectively.

1.2 RESEARCH OBJECTIVES

1.2.1 Long-term objective

This study is part of a long-term program to develop a multi-probe fluorescence-based detection assay for the BoNTA. Fluorescence Polarization as a fast and reliable detection assay will be used for BoNTA detection. In addition, new fluorescent recognition agents for BoNTA will be identified through the application of computer-assisted drug design techniques.

1.2.2 Specific aims of this study

1) Use computer-assisted drug design techniques (CADD) to identify recognition agents for BoNTA detection, with emphasis on fluorescent ligands.

2) Assess Fluorescence Polarization as assay platform for the development of a fast and reliable detection assay for BoNTA using newly identified fluorescent recognition agents.

1.3 REFERENCES

1. Fujinaga, Y., Interaction of botulinum toxin with the epithelial barrier. *Journal of biomedicine & biotechnology* **2010,***2010*, 974943.

2. Cai, S.; Singh, B. R.; Sharma, S., Botulism diagnostics: from clinical symptoms to in vitro assays. *Critical reviews in microbiology* **2007,***33* (2), 109-25.

3. Arnon, S. S.; Schechter, R.; Inglesby, T. V.; Henderson, D. A.; Bartlett, J. G.; Ascher, M. S.; Eitzen, E.; Fine, A. D.; Hauer, J.; Layton, M.; Lillibridge, S.; Osterholm, M. T.; O'Toole, T.; Parker, G.; Perl, T. M.; Russell, P. K.; Swerdlow, D. L.; Tonat, K., Botulinum toxin as a biological weapon: medical and public health management. *JAMA : the journal of the American Medical Association* **2001,***285* (8), 1059-70.

4. (a) Sharma, S. K.; Ferreira, J. L.; Eblen, B. S.; Whiting, R. C., Detection of type A, B, E, and F Clostridium botulinum neurotoxins in foods by using an amplified enzyme-linked immunosorbent assay with digoxigenin-labeled antibodies. *Applied and environmental microbiology* **2006,***72* (2), 1231-8;(b) Hasegawa, K.; Watanabe, T.; Suzuki, T.; Yamano, A.; Oikawa, T.; Sato, Y.; Kouguchi, H.; Yoneyama, T.; Niwa, K.; Ikeda, T.; Ohyama, T., A novel subunit structure of Clostridium botulinum serotype D toxin complex with three extended arms. *The Journal of biological chemistry* **2007,***282* (34), 24777-83.

5. Dhaked, R. K.; Singh, M. K.; Singh, P.; Gupta, P., Botulinum toxin: bioweapon & magic drug. *The Indian journal of medical research* **2010,***132* (5), 489-503.

6. Whelchel, D. D.; Brehmer, T. M.; Brooks, P. M.; Darragh, N.; Coffield, J. A., Molecular targets of botulinum toxin at the mammalian neuromuscular junction. *Movement disorders : official journal of the Movement Disorder Society* **2004,***19 Suppl 8*, S7-S16.

7. Boldt, G. E.; Kennedy, J. P.; Hixon, M. S.; McAllister, L. A.; Barbieri, J. T.; Tzipori, S.; Janda, K. D., Synthesis, characterization and development of a high-throughput methodology for the discovery of botulinum neurotoxin a inhibitors. *Journal of combinatorial chemistry* **2006,***8* (4), 513-21.

8. Lindstrom, M.; Korkeala, H., Laboratory diagnostics of botulism. *Clinical microbiology reviews* **2006,***19* (2), 298-314.

9. Cai, S.; Lindo, P.; Park, J. B.; Vasa, K.; Singh, B. R., The identification and biochemical characterization of drug-like compounds that inhibit botulinum neurotoxin serotype A endopeptidase activity. *Toxicon : official journal of the International Society on Toxinology* **2010,***55* (4), 818-26.

10. Cherington, Clinical spectrum of botulism. *muscle & nerve* **1998,***21* (6), 701-710.

11. Sharma, Evaluation of lateral-flow Clostridium botulinum neurotoxin detection kits for food analysis. **2005**.

12. Wu, Detection of Clostridium botulinum neurotoxin type A using immuno-PCR. *Letter in applied microbiology* **2001,***32* (5), 321-325.

13. (a) Smith, Fluorescence polarization immunoassays and related methods for simple, high-throughput screening of small molecules. *Analytical and bioanalytical chemistry* **2008,***391* (5);(b) Gagne, Use of fluorescence polarization detection for the measurement of fluopeptidetm binding to G protein-coupled receptors. *Receptor and signal transduction research* **2002,***22* (1-4);(c) Chun, H. S.; Choi, E. H.; Chang, H. J.; Choi, S. W.; Eremin, S. A., A fluorescence polarization immunoassay for the detection of zearalenone in corn. *Analytica chimica acta* **2009,***639* (1-2), 83-9.

14. (a) Nielsen, K.; Gall, D., Fluorescence polarization assay for the diagnosis of brucellosis: a review. *Journal of immunoassay & immunochemistry* **2001,***22* (3), 183-201;(b) Smith, D. S.; Eremin, S. A., Fluorescence polarization immunoassays and related methods for simple, high-throughput screening of small molecules. *Anal Bioanal Chem* **2008,***391* (5), 1499-507;(c) Parker, G. J.; Law, T. L.; Lenoch, F. J.; Bolger, R. E., Development of high throughput screening assays using fluorescence polarization: nuclear receptor-ligand-binding and kinase/phosphatase assays. *Journal of biomolecular screening* **2000,***5* (2), 77-88.

15. Jameson, D. M.; Croney, J. C., Fluorescence polarization: past, present and future. *Combinatorial chemistry & high throughput screening* **2003,***6* (3), 167-73.

16. Lee, P. H.; Bevis, D. J., Development of a homogeneous high throughput fluorescence polarization assay for G protein-coupled receptor binding. *Journal of biomolecular screening* **2000,***5* (6), 415-19.

17. Goddard, J. P.; Reymond, J. L., Recent advances in enzyme assays. *Trends in biotechnology* **2004,***22* (7), 363-70.

18. Feltrup, T. M.; Singh, B. R., Development of a fluorescence internal quenching correction factor to correct botulinum neurotoxin type A endopeptidase kinetics using SNAPtide. *Analytical chemistry* **2012,***84* (24), 10549-53.

19. Joshi, S. G., Detection of biologically active botulinum neurotoxin--A in serum using high-throughput FRET-assay. *Journal of pharmacological and toxicological methods* **2012,***65* (1), 8-12.

20. Sun, S.; Almaden, J.; Carlson, T. J.; Barker, J.; Gehring, M. R., Assay development and data analysis of receptor-ligand binding based on scintillation proximity assay. *Metabolic engineering* **2005,***7* (1), 38-44.

21. (a) de Jong, L. A.; Uges, D. R.; Franke, J. P.; Bischoff, R., Receptor-ligand binding assays: technologies and applications. *Journal of chromatography. B, Analytical technologies in the biomedical and life sciences* **2005,***829* (1-2), 1-25;(b) Glickman, J. F.; Schmid, A.; Ferrand, S., Scintillation proximity assays in high-throughput screening. *Assay and drug development technologies* **2008,***6* (3), 433-55.

22. Schmidt, T.; Bergner, A.; Schwede, T., Modelling three-dimensional protein structures for applications in drug design. *Drug discovery today* **2013**.

23. (a) Yang, S. Y., Pharmacophore modeling and applications in drug discovery: challenges and recent advances. *Drug discovery today* **2010,***15* (11-12), 444-50;(b) Wolber, G.; Seidel, T.; Bendix, F.; Langer, T., Molecule-pharmacophore superpositioning and pattern matching in computational drug design. *Drug discovery today* **2008,***13* (1-2), 23-9.

24. Pirhadi, S.; Shiri, F.; Ghasemi, J. B., Methods and applications of structure based pharmacophores in drug discovery. *Current topics in medicinal chemistry* **2013,***13* (9), 1036-47.

25. (a) Ma, D. L.; Chan, D. S.; Leung, C. H., Drug repositioning by structure-based virtual screening. *Chemical Society reviews* **2013,***42* (5), 2130-41;(b) Tan, L.; Batista, J.; Bajorath, J., Computational methodologies for compound database searching that utilize experimental protein-ligand interaction information. *Chemical biology & drug design* **2010,***76* (3), 191-200;(c) Bajorath, J., Computational analysis of ligand relationships within target families. *Current opinion in chemical biology* **2008,***12* (3), 352-8.

26. Meng, X. Y.; Zhang, H. X.; Mezei, M.; Cui, M., Molecular docking: a powerful approach for structure-based drug discovery. *Current computer-aided drug design* **2011,***7* (2), 146-57.

27. Kumar, V.; Chandra, S.; Siddiqi, M. I., Recent Advances in The Development of Antiviral Agents Using Computer-Aided Structure Based Approaches. *Current pharmaceutical design* **2013**.

28. Ramjan, Z. H.; Raheja, A.; Floriano, W. B., A cluster-aware graphical user interface for a virtual ligand screening tool. Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference 2008, 2008, 4102-5. 2.

29. Sitter, T.; Willick, D. L.; Floriano, W. B., Computational Drug Screening in the Cloud Using HierVLS/PSVLS. The 2013 World Congress in Computer Science Compouter Engineering and Applied Computing. Las Vegas, Nevada, USA, July 22-25 2013.

2 CHAPTER TWO: STRUCTURAL AND SEQUENCE ANALYSIS OF BoNT SEROTYPES (A-G) AND BoNTA SUBTYPES

2.1 INTRODUCTION

Structural analysis of different subtypes of BoNTA will improve our understanding of their structural features. Additionally, this study can enhance our knowledge of different BoNT serotypes (A-G) which is useful in the design of subtype-specific assays. The goal is to identify critical regions that may differentiate BoNT subtypes for detection methods. In this project, bioinformatics tools and database information were utilized to determine regions of high and low similarity between different subtypes of BoNTA (A1, A2, A3 and A5), and also different serotypes of BoNT (A, B, C, D, E, F, and G).

2.2 METHODS AND PROCEDURES

2.2.1 Sequence and structural analyses of BoNTA subtypes

2.2.1.1 *boNTA1* gene

The information about *boNTA1* gene was obtained from the National Center For Biotechnology Information (NCBI) gene bank [1]. Three BoNTs were found using the NCBI and searching for *boNTA1* gene. These included (1) botulinum neurotoxin type A1 from *Clostridium botulinum A str. Hall*, (2) botulinum neurotoxin type A1 from *Clostridium botulinum F str. Langeland* and (3) botulinum neurotoxin type A1 from *Clostridium botulinum A str. (ATCC 19397)*. The botulinum neurotoxin type A1 from *Clostridium botulinum A str. Hall* was chosen

for further analysis [2]. The information about *boNTA* gene was found using the "botulinum neurotoxin type A1 [*Clostridium botulinum A str.Hall*]" link. The gene product information was obtained from "UniportKB/Swiss-Prot" link under "mRNA and Protein" [1].

2.2.1.2 The amino acid sequence of boNTA1

The amino acid sequence of BoNTA (accession number: A5HZZ9) was obtained from "UniportKB" database of the ExPASy (Export Protein Analysis System) proteomics server website. The protein sequence is shown in Figure 2.1.

2.2.1.3 BLAST search

Database similarity search with the BoNTA amino acid sequence was performed using BLAST search. The goal of this step was to search for similarity between BoNTA sequence and other sequences in the database to find homologous proteins. The UniportKB database of the ExPASy website was used for BLAST search. The parameters for BLAST search were as follow: substitution matrix: BLOSUM-62, threshold: 10, filtering: filter low complexity region, gap: was considered, and hits: 100.

2.2.1.4 Multiple sequence alignment

To identify regions of similarity and diversion between different subtypes of BoNT, a multiple sequence alignment was performed. The amino acid sequences of BoNTA were saved into a single file in fasta format, readable by the alignment program, Clustal[3]. The multiple sequence alignment was run using Clustal x with the following parameters: substitution matrix: BLOSUM30, gap Opening penalty: 10, and gap extension penalty: 0.10.

2.2.1.5 Identification of conserved motifs

Motifs, domains and pattern are used to classify proteins into functional families. A pattern is a small part of sequence with distinct amino acids at specific position. This region is conserved within a family of proteins. The PROSIT, PRINTS, and PFAM were used to identified pattern, motifs and domains of BoNTA [4].

2.2.1.6 3D structure of BoNTA

To analyze BoNTA 3D structures, first the complete structure of BoNTA was obtained from the protein data bank. 2NYY and 3BTA were two choices of complete structure of BoNTA. In this study 3BTA structure was selected for this structural analysis. The program VMD[5] was used to visualize the BoNTA 3D structure and identify critical regions of the protein [6]. Also, variable residues obtained by multiple alignment of the different subtypes of BoNTA (Table 2.1) were visualized within the structure in order to better examine the variable regions.

2.2.1.7 Quality of the protein structure

The Procheck (www.ebi.ac.uk) and Whatcheck (www.ebi.ac.uk) programs were used to analyze the structure 3BTA in term of chirality, hydrogen bonds, torsion angles, missing atoms and segments [7]. This was performed to assess the quality of the structure.

2.2.2 Sequence and structural analyses of BoNT serotypes

2.2.2.1 The amino acid sequence of BoNT serotypes

The amino acid sequences of BoNT serotypes (A-G) were collected from "UniportKB" database of the ExPASy (Export Protein Analysis System) proteomics server website to use as input for the multiple alignment [8].

2.2.2.2 BLAST search

The BoNT serotype amino acid sequences were obtained through a database similarity search using BLAST. This step is necessary to identify similarities between BoNT serotype sequences. The BLAST search result contains two parameters to determine the significance of the hits including score and expectation value (e-value). Higher score indicates higher similarity of the amino acid sequences and very small value of e-value (≤0) confirms that blast search is statistically significant. The UniportKB database of the ExPASy website (www.expasy.org) was used for BLAST search. The parameters for BLAST search were as follow: substitution matrix: BLOSUM-62, threshold: 10, filtering: filter low complexity region, gap: was considered, and hits: 100.

2.2.2.3 Multiple sequence alignment

A multiple sequence alignment was performed to determine regions of high and low similarity between the different BoNT serotypes. The multiple alignment was also used to highlight important regions for discrimination between BoNT serotypes. The amino acid sequences of the different BoNT serotypes were saved into a single file in fasta format readable by Clustal. The multiple sequence alignment was run using Clustal x with the following parameters: substitution matrix: BLOSUM30, gap Opening penalty: 10, and gap extension penalty: 0.10.

2.2.2.4 Identification of conserved motifs

The PRINTS database was used to find conserved motifs between BoNT serotypes [9]. Also, PFAM and PROSITE were searched to determine domains and patterns within the sequence of BoNT serotypes [4].

2.2.2.5 Visualization of variable residues in BoNT serotypes

As mentioned in the section 2.1.6. (3D structure of BoNTA), the 3D structure of BoNTA and variable residues obtained by multiple alignment (Table 2.5) were used to visualize the variable regions of the different serotypes of BoNT.

2.3 RESULTS &DISCUSSION

2.3.1 Sequence and structural analyses of BoNTA subtypes

2.3.1.1 *boNTA1* gene

boNTA1 (accession number: A5HZZ9) encodes for botulinum neurotoxin type A (BoNTA). Botulinum neurotoxin type A binds to the largest luminal loop of the secretory vesicle protein SV2 (SV2A, SV2B, and SV2C) and then enters the cytoplasm. The heavy chain is responsible for binding to nerve cell receptors and transferring the light chain into cytoplasm. The light chain is responsible for cleavage of SNARE proteins (including SNAP25, synaptobrevins, and syntaxin) at 197-Gln-|-Arg-198 location which blocks neurotransmitter release. The light chain's catalytic mechanism is a metalloprotease which needs a zinc for its catalytic activity [10]. BoNTA belongs to the peptidase M27 family and contains 1296 amino acids. The light chain starts from amino acid number 2 to 448 (447 AA length) and the heavy chain starts from amino acid 449 to 1296 (848 AA length). BoNTA contains two transmembrane regions and each region consists of 21 AA. One of the transmembrane regions begins from amino acid 627 to 647 and the other one starts from amino acid 656 to 676 (Figure 2.2). The amino acid number 224 is in the active site and amino acids numbers 223, 227, and 262 are metal binding sites (Figure 2.3). The two disulfide bonds which are connecting light chain to heavy chain are located between amino acid 430↔454 and 1235↔1280 (Figure 2.4). Based on

experimental data, the amino acid 1218 (S→Y) is a sequence conflict and a fully active BoNT is extends at least up to 1218 amino acids[10].

2.3.1.2 The amino acid sequence of BoNTA1

The amino acid sequence of BoNTA1 in fasta format is shown in Figure 2.1.

```
>sp|A5HZZ9|BXA1_CLOBH Botulinum neurotoxin type A OS=Clostridium botulinum
     (strain Hall / ATCC 3502 / NCTC 13319 / Type A) GN=botA PE=1 SV=1

        MPFVNKQFNYKDPVNGVDIAYIKIPNAGQMQPVKAFKIHNKIWVIPERDTFTNPEEGDLN
        PPPEAKQVPVSYYDSTYLSTDNEKDNYLKGVTKLFERIYSTDLGRMLLTSIVRGIPFWGG
        STIDTELKVIDTNCINVIQPDGSYRSEELNLVIIGPSADIIQFECKSFGHEVLNLTRNGY
        GSTQYIRFSPDFTFGFEESLEVDTNPLLGAGKFATDPAVTLAHELIHAGHRLYGIAINPN
        RVFKVNTNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQENEFRLYYYNKFKDIASTLNKA
        KSIVGTTASLQYMKNVFKEKYLLSEDTSGKFSVDKLKFDKLYKMLTEIYTEDNFVKFFKV
        LNRKTYLNFDKAVFKINIVPKVNYTIYDGFNLRNTNLAANFNGQNTEINNMNFTKLKNFT
        GLFEFYKLLCVRGIITSKTKSLDKGYNKALNDLCIKVNNWDLFFSPSEDNFTNDLNKGEE
        ITSDTNIEAAEENISLDLIQQYYLTFNFDNEPENISIENLSSDIIGQLELMPNIERFPNG
        KKYELDKYTMFHYLRAQEFEHGKSRIALTNSVNEALLNPSRVYTFFSSDYVKKVNKATEA
        AMFLGWVEQLVYDFTDETSEVSTTDKIADITIIIPYIGPALNIGNMLYKDDFVGALIFSG
        AVILLEFIPEIAIPVLGTFALVSYIANKVLTVQTIDNALSKRNEKWDEVYKYIVTNWLAK
        VNTQIDLIRKKMKEALENQAEATKAIINYQYNQYTEEEKNNINFNIDDLSSKLNESINKA
        MININKFLNQCSVSYLMNSMIPYGVKRLEDFDASLKDALLKYIYDNRGTLIGQVDRLKDK
        VNNTLSTDIPFQLSKYVDNQRLLSTFTEYIKNIINTSILNLRYESNHLIDLSRYASKINI
        GSKVNFDPIDKNQIQLFNLESSKIEVILKNAIVYNSMYENFSTSFWIRIPKYFNSISLNN
        EYTIINCMENNSGWKVSLNYGEIIWTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTIT
        NNRLNNSKIYINGRLIDQKPISNLGNIHASNNIMFKLDGCRDTHRYIWIKYFNLFDKELN
        EKEIKDLYDNQSNSGILKDFWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRGYMYLKGPR
        GSVMTTNIYLNSSLYRGTKFIIKKYASGNKDNIVRNNDRVYINVVVKNKEYRLATNASQA
        GVEKILSALEIPDVGNLSQVVVMKSKNDQGITNKCKMNLQDNNGNDIGFIGFHQFNNIAK
         LVASNWYNRQIERSSRTLGCSWEFIPVDDGWGERPL
```

Figure 2.1 BoNTA amino acid sequence in fasta format [10].

2.3.1.3 BLAST search

Amino acid sequence of BoNTA1 (accession number: A5HZZ9) was used as query for a BLAST search. Based on the BLAST result about 25% of the hits had more than 90% similarity and also more than 83% identity. The other 75% showed approximately less than 58% similarity

with 40% identity. The E-values for all the sequences were equal to 0.0, thus the BLAST search was statistically significant. The score values for the 25% (with high identity) were between 6710-5700. The10 hits with the similarity more than 90% were chosen for further analyses. Table 2.1 summarizes the information related to the selected 10 hits.

2.3.1.4 Multiple sequence alignment

Multiple sequence alignment of the 10 selected hits was performed using Clustal. Based on the multiple sequence alignment, the selected 10 sequences showed high similarity as expected. The variable amino acids, weakly and strongly conserved residues as well as strongly conserved regions are summarizes in the Table 2.2. The critical regions of the BoNTA sequences were found based on the information obtained from sequence annotation feature of the ExPASy proteomics server website [10] and the multiple alignment. As shown in the Figure 2.2, the transmembrane regions of the selected 10 BoNTA amino acid sequences are fully conserved. Each conserved helical transmembrane region contains 21 amino acids. The first region starts at amino acid 627 and ends at 647 and the second one starts at amino acid 656 and ends at 676. The active site (Glu224) and zinc binding sites (His223, His227 and Glu262) in the 10 selected BoNTA were also fully conserved (Figure 2.3) which is in accordance to studies by Arndt (2006) [11]. As shown in the Figure 2.4, the four amino acid cysteines (430, 454, 1235 and 1280) involving in disulfide bonds are conserved in the all 10 selected amino acid sequences.

Table 2.1 The information of the 10 selected protein sequences with high similarity to BoNTA (>90%) are summarized.

	Accession	Description	Length	Organism	Family	Score	E-value	% Identity	% Similarity
1	A7FS63	Bontoxilysin A (EC 3.4.24.69)	1296	Bontoxilysin A (EC 3.4.24.69)	Clostridiaceae	6710	0.0	99%	99%
2	Q7B8V4	BoNTA (Bont/A1) (Botulinum neurotoxin type A) (Neurotoxin A) (Neurotoxin BoNT)	1296	Clostridium botulinum	Clostridiaceae	6710	0.0	99%	99%
3	A5HZZ9	Botulinum neurotoxin type A (BoNTA) (EC 3.4.24.69) (Bontoxilysin-A) (BOTOX) [Cleaved into: Botulinum neurotoxin A light chain; Botulinum neurotoxin A heavy chain]	1296	Clostridium botulinum (strain Hall / ATCC 3502 / NCTC 13319 / Type A)	Clostridiaceae	6710	0.0	99%	99%
4	C6K838	BoNTA1	1296	Clostridium botulinum	Clostridiaceae	6706	0.0	99%	99%
5	P10845	Botulinum neurotoxin type A (BoNTA) (EC 3.4.24.69) (Bontoxilysin-A) (BOTOX) [Cleaved into: Botulinum neurotoxin A light chain; Botulinum neurotoxin A heavy chain]	1296	Clostridium botulinum	Clostridiaceae	6706	0.0	99%	99%
6	B1Q6N6	Botulinum neurotoxin type A1, BoNTA1 (EC 3.4.24.69)	1296	Clostridium botulinum NCTC 2916	Clostridiaceae	6698	0.0	98%	98%
7	C7BEA8	Botulinum neurotoxin type A5	1296	Clostridium botulinum	Clostridiaceae	6540	0.0	96%	97%
8	Q58GH1	BoNTA2	1296	Clostridium botulinum	Clostridiaceae	6101	0.0	89%	94%
9	C3KS13	Botulinum neurotoxin type BvA4, BoNTBvA4	1296	Clostridium botulinum	Clostridiaceae	6009	0.0	88%	92%
10	D3IV24	BoNTA3	1292	Clostridium botulinum	Clostridiaceae	5722	0.0	83%	90%

Table 2.2 Summary of the information obtained from multiple alignment of the 10 selected BoNTA sequences.

Variable amino acids	2, 14, 41, 50, 57, 92, 109, 110, 140, 154, 163, 169,215, 229, 264, 269, 278, (283 -286), 295, 296, 319, 332, 336, 337, 348, 352, 357, 382, 387, (397– 401), 406, 411, 437, 478, 496, 515, 530, 567, 581, 593, 601, 602, 607, 624, 648, 654, 663, 671, 713, 775, 779, 832, 874, 882,899, 900, 905, 993, 994, 1006, 1156, 1232, 1254, 1259, 1272, 1294
Weakly conserved	11, 28, 121, 122, 141, 142, 152, 166, 170, 194, 212, 239, 309, 319, 339, 347, 381, 393, 394, 410, 436, 505, 572, 580, 582, 589, 598, 605, 619, 650, 661, 707, 767, 804, 818, 880, 885, 887, 895, 902, 908, 923, 954, 955, 957, 990, 1110, 1141, 1142, 1208, 1295
Strongly conserved	29, 40, 42, 70, 76, 96, 113, 137, 139, 159, 162, 164, 171, 173, 220, 243, 246, 252, 260, 269, 271, 272, (279-281),(288-293), 299, 301, 303, 308, 316, 322, 328, 329, 331, 334, 335, 340, 341, 343, 344, 349, 353, 356, 361, 375, 388, 395, 344, 352, 376, 415, 444, 452, 476, 483, 507, 546, 556, 573, 578, 591, 592, 594, 604, 609, 622, 634, 635, 647, 651, 656, 657, 659, 664, 667, 668, 672, 673, 676, 681, 685, 686, 696, 726, 730, 734, 735, 744, 789, 809, 815, 816, 831, 835, 840, 847, (858-861), 876, 881, 884, 886, 896, 897, 903, 906, 907, 915, 917, 926, 930, 948, 967, 991, 992, 1003, 1117, 1128, 1143, 1144, 1153, 1170, 1228, 1250, 1255, 1256, 1271, 1273, 1274, 1278
Conserved regions	1,(3-10), (12-13), (15-26), 28, (30-39), (43-50), (51-56), (58-69), (71-75), (77-91), 93, 95, (97-101), (103-108), 111, 112, (114-120), (123-136), 138, (143-151), 153, (155-158), 160, 161, 165, 167, 168, 172, (174-193, (195-211), 213, 214, (216-219), (221-228), (230-238), (240-242), 244, 245, (247-253), (255-259), (261-263), (265-267), 270,(273-277), 282, 287, 294, 297, 298, (300-301), 303, (305-307), (310-315), 317, 320, 321, (323-326), 329, 330, 333, 338, 342, 345, 346, 350, 351, 354, 355, (358-360), (362-380), (383-386), (389-392), 396, (501-504), 506, (508-514), (508-514), (516-529), (531-545), (547-559), 561, (564-566), (568-571), (574-577), 579, 623, (625-632), (635-646), 649, (652-653), 655, 658, 660, 662, 665, 666, 669, 670, 674, 675, (677-680), (682-684), (687-695), (696-706), (708-712), (714-719), (721-725), (727-729), (731-733), (736-743), (745-766), (768-774),(776-778), (780-788), (790-803), (805-808), (810-814), 817, (819-830), 833, 834, (836-839), (841-846), (848-857), (862-873), 875, (877-879), 883, (888-894), 898, (909-914), 916, (918-922), 924, 925, (927-929), (931-947), (949-953), 956, (958-967), (969-989), (995-1004), (1007-1014), 1016, (1018-1023), (1027-1051), (1053-1062), (1065-1067), (1069-1079), (1081-1089), (1091-1093), (1095-1102), (1104-1109), (1111-1116), (1118-1127), (1129-1140), (1145-1152), 1154, 1155, (1157-1169), (1171-1207), (1209-1226), (1228-1231), (1233-1249), (1251-1253), (1257-1258), (1260-1270), (1275-1277), (1279-1291), 1293, 1296

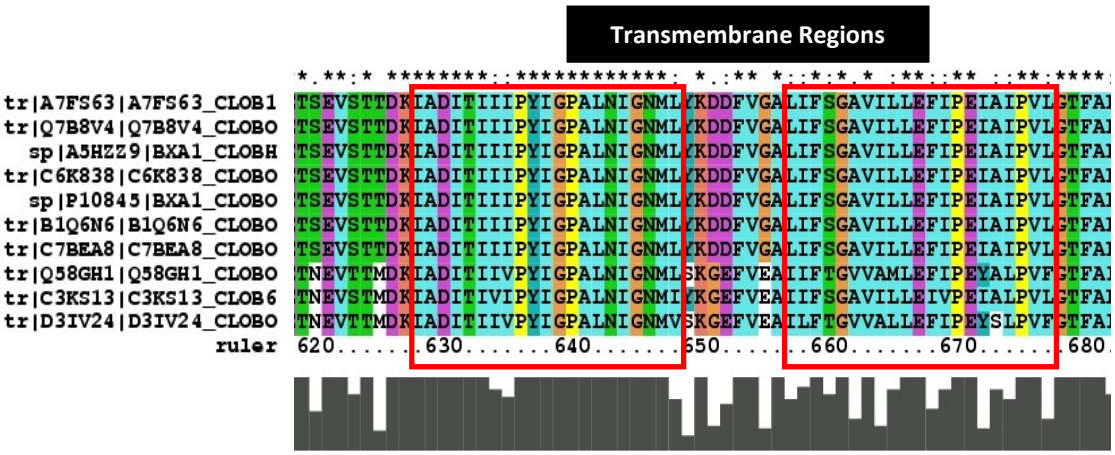


Figure 2.2 Multiple alignment of the 10 selected BoNTA subtypes. The transmembrane regions shown in the red boxes are fully conserved in the selected 10 BoNTA amino acid sequences.

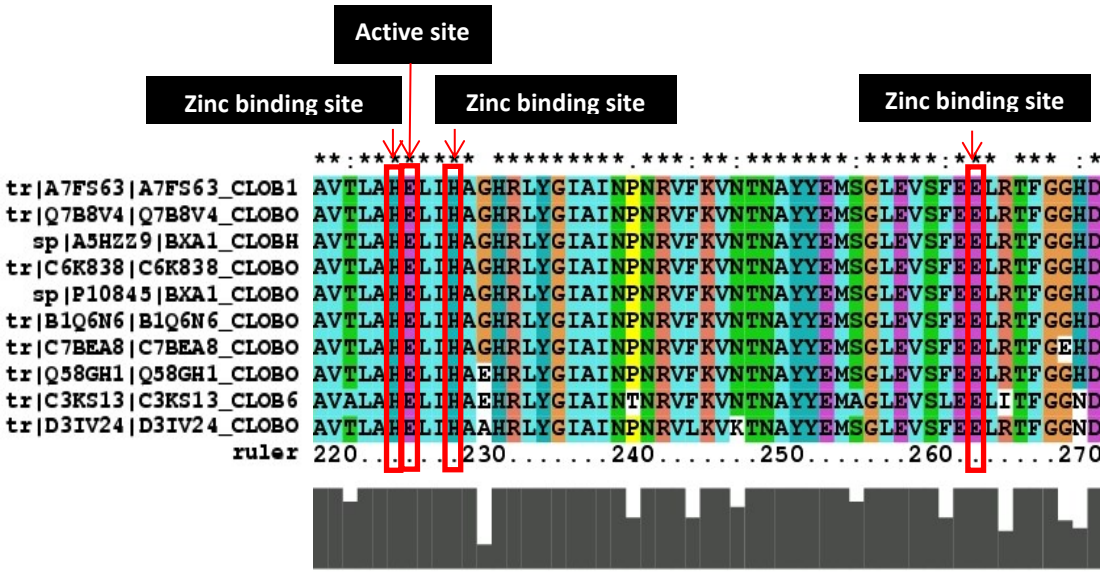


Figure 2.3 Multiple sequence alignment of the 10 selected BoNTA subtypes. The active site and zinc binding sites shown by the arrows are fully conserved in the selected 10 BoNTA amino acid sequences.

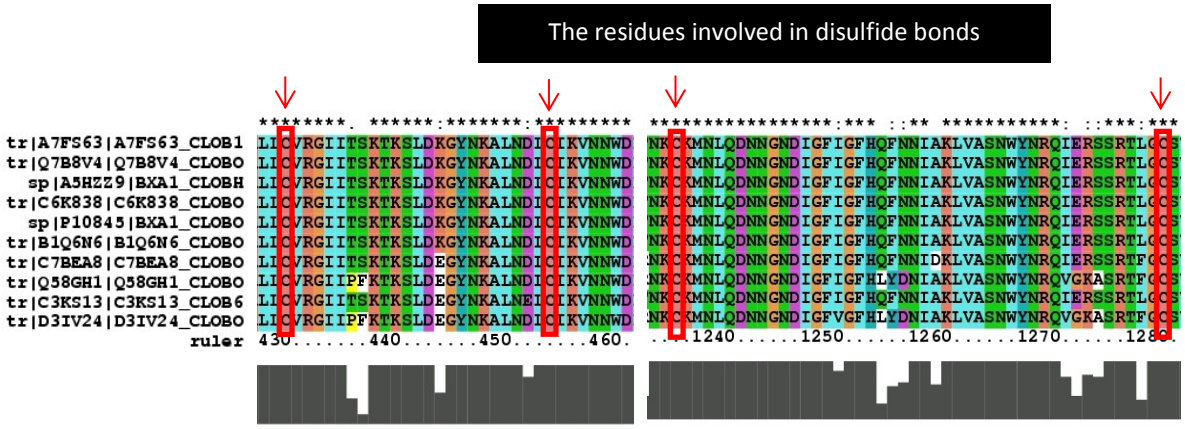


Figure 2.4 Multiple alignment of the 10 selected BoNTA subtypes. The conserved amino acids (cysteine) involved in disulfide bonds are shown by the arrows.

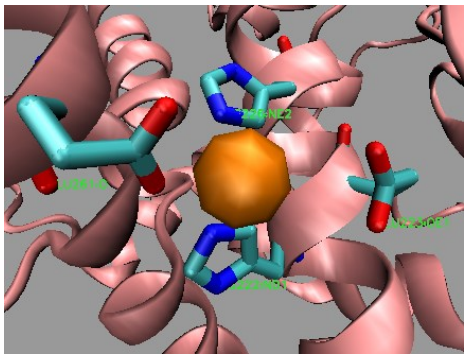
2.3.1.5 Identification of conserved motifs

The BoNTA was characterized by the zinc protease motif pattern using PROSITE. The results showed that BoNTA belongs to peptidase M27 family and zinc-binding region signature is as follow **[**G S T A L I V N] - { P C H R } - { K N D } - H - E - [L I V M F Y W] - { D E H R K P } - H - { E K P C } - [L I V M F Y W G S P Q] . This bracket [] in pattern sequence means a protein sequence required to have any amino acids at this specific position except the amino acids listed inside the bracket [] to be a member of this family. This bracket {} of pattern sequence means either one of the amino acids listed inside the bracket should be available at this specific position to consider a protein as a member of this family. PRINTS and PFAM databases were searched to identify motifs and domains within BoNTA sequence. The number of motifs and domains were six (1-6) and four, respectively. The four BoNT domains are peptidase_M27, transmembrane, N-terminal and C-terminal domains [9, 12].

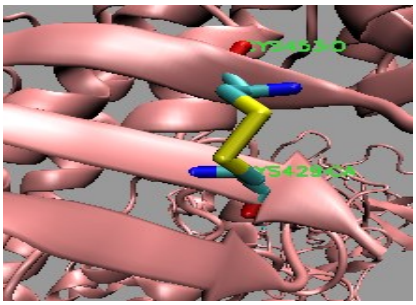
2.3.1.6 3D structure of BoNTA

The 3D structure of BoNTA was visualized using 3BTA pdb file in VMD and presented in the Figure 2.5. The amino acids in the active site involved in zinc coordination are shown in the Figure 2.5A. Additionally the amino acids involved in disulfide bonds are presented in Figure 2.5B and C. The resolution of 3BTA crystallography structure was 3.20 Å. BoNTA consists of 19 α-helices and 54 β-strand regions which were obtained from analyzing the 3D structure using VMD. The variable residues between BoNTA subtypes obtained based on the multiple alignment results (Table 2.2) visualized in the Figure 2.6.

A)



B) **C)**



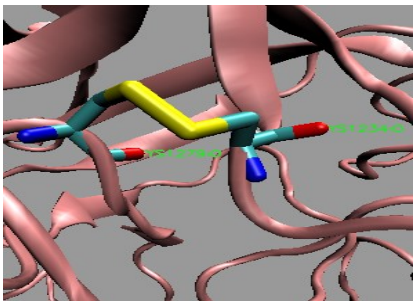


Figure 2.5 The 3D structure of BoNTA. A, represents the amino acids in the active site involved in zinc coordination (sphere is representing the zinc atom). B and C, represent the four amino acids involved in disulfide bonds.

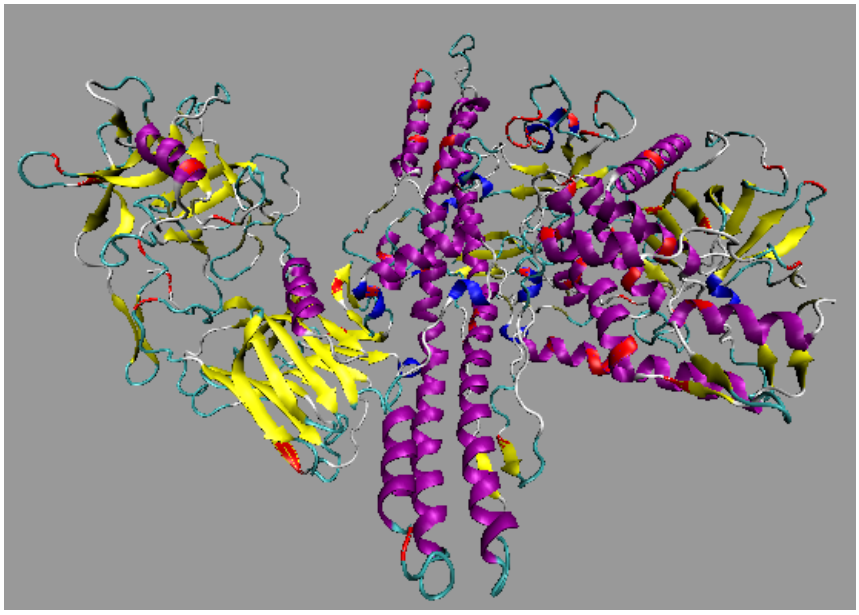


Figure 2.6 Variable residues identified between BoNTA subtypes. The red colors represent the variable residues found in BoNTA subtypes using multiple sequence alignment. The yellow and purple colors represent β-strands and α-helices, respectively.

2.3.1.7 Quality of the protein structure

The Psi/Phi angles for the 3BTA structure are concentrated in "most favoured regions" of the Ramachandran plot. The number of Glycine and Proline in the 3BTA structure is 62 and 38, respectively [13]. Based on Whatcheck analysis report, there are two missing residues in the 3BTA structure which are Asparagine 432 and Arginine 431. Increasing the number of missing atoms will decrease the sensitivity of the structure quality check. The number of reported unsatisfied hydrogen bond donors and unsatisfied hydrogen bond acceptors in the structure were 25 and 4, respectively. These hydrogen atoms are not involved in H-bond network. Less unsatisfied H-bond represents good quality. The number of residue contains abnormal torsion angles were 25

and also all 3BTA protein atoms have proper chirality [14]. According to the previous information, the overall quality of 3BTA structure was reasonable.

2.3.2 Sequence and structural analyses of BoNT serotypes

2.3.2.1 The amino acid sequence of BoNT serotypes

The critical regions of the BoNT serotype sequences were found based on the obtained information from the sequence annotation feature of the ExPASy proteomics server website. These features are including the length of the light and heavy chains, transmembrane regions, active sites, metal binding sites and the residues involved in disulfide bonds. Table 2.3 summarizes these features of the BoNT serotypes.

2.3.2.2 BLAST search

The amino acid sequences for the different BoNT serotypes A-G were obtained through a BLAST search using amino acid sequence of BoNTA (accession number: P10845) as the query. The BLAST search was statistically significant because the e-value for each pair was equal to 0.0. The score values were between 6710-1700. The similarity and identity percentages of sequences were more than 50% and 31%, respectively. The BLAST search of the different serotypes of BoNT (A-G) information was provided in the Table 2.4.

Table 2.3 The BoNT serotypes sequences annotation features.

Accession # And Name	P10845 BoNTA	P30996 BoNTF	Q60393 BoNTG	Q00496 BoNTE	P10844 BoNTB	P19321 BoNTD	Q5DW55 BoNTC
Light Chain	2-448	1-436	2-442	2-422	2-441	1-442	-
Heavy chain	449-1296	437-1274	443-1297	423-1251	442-1291	443-1276	-
Transmembrane Region	627-647 656-676	-	-	638-658	653-673	-	-
Active site	224	228	231	213	231	230	-
Metal binding	223, 227, 262	227, 231	230, 234	212, 216	230, 234	229, 233, 269	-
Disulfide bond	430↔454 1235↔1280	429↔445	436↔450	412↔426	437↔446	437↔450	-
Neutral variant	27 V↔A	-	-	-	-	15-16 ND→PV 17-18 ND→LQ 452 K→Q, 457 R→F 457 R→T, 462 A→D 489 K→N, 644 N→K 1122Q→R	-
Mutagenesis	262 E→A 266 F→A 366 Y→A (decreases enzymatic activity)	-	-	-	-	-	-
Sequence conflict	2 P→Q 408 E→P 876 T→L 892 S→K	-	-	177 R→G 198 C→S 340 R→A 773 I→L 177 R→G 963-964 FE→LQ 967 R→A 1195 N→NN	30 T→M 218 R→G 225A →S 464 S→R	-	-
length	1296	1274	1297	1251	1291	1276	1280
References	[15]	[16]	[17]	[18]	[19]	[20]	[21]

Table 2.4 BLAST search results for different BoNT serotypes.

	Accession	Description	Length	Organism	Family	Score	E-value	% Identity	% Similarity
1	P10845	Botulinum neurotoxin type A (BoNTA) (EC 3.4.24.69) (Bontoxilysin-A) (BOTOX) [Cleaved into: Botulinum neurotoxin A light chain; Botulinum neurotoxin A heavy chain]	1296	Clostridium botulinum	Clostridiaceae	6710	0.0	99%	99%
2	P30996	Botulinum neurotoxin type F (BoNTF) (EC 3.4.24.69) (Bontoxilysin-F) [Cleaved into: Botulinum neurotoxin F light chain; Botulinum neurotoxin F heavy chain]	1274	Clostridium botulinum	Clostridiaceae	2171	0.0	38%	57%
3	Q60393	Botulinum neurotoxin type G (BoNTG) (EC 3.4.24.69) (Bontoxilysin-G) [Cleaved into: Botulinum neurotoxin G light chain; Botulinum neurotoxin G heavy chain]	1297	Clostridium botulinum	Clostridiaceae	2167	0.0	38%	55%
4	Q00496	Botulinum neurotoxin type E (BoNTE) (EC 3.4.24.69) (Bontoxilysin-E) [Cleaved into: Botulinum neurotoxin E light chain; Botulinum neurotoxin E heavy chain]	1251	Clostridium botulinum	Clostridiaceae	2162	0.0	38%	57%
5	P10844	Botulinum neurotoxin type B (BoNTB) (EC 3.4.24.69) (Bontoxilysin-B) [Cleaved into: Botulinum neurotoxin B light chain; Botulinum neurotoxin B heavy chain]	1291	Clostridium botulinum	Clostridiaceae	2145	0.0	38%	55%
6	P19321	Botulinum neurotoxin type D (BoNTD) (EC 3.4.24.69) (Bontoxilysin-D) [Cleaved into: Botulinum neurotoxin D light chain; Botulinum neurotoxin D heavy chain]	1276	Clostridium botulinum	Clostridiaceae	1764	0.0	32%	52%
7	Q5DW55	Type C neurotoxin	1280	Clostridium botulinum	Clostridiaceae	1700	0.0	31%	50%

2.3.2.3 Multiple sequence alignment

Multiple sequence alignment of the different BoNT serotypes was performed by Clustal[3]. Based on the multiple sequence alignment, the similarity between BoNT serotypes sequences were less than the similarity between BoNTA subtypes, as expected. The variable amino acids, weakly and strongly conserved residues as well as strongly conserved regions are summarized in the Table 2.5. The critical regions for differentiation of BoNT serotypes sequences were found based on the multiple alignment and ExPASy website information. The transmemberane regions of the BoNTA were compared to the other BoNT serotypes in Figure 2.7. The result indicates that the regions are not fully conserved between the different serotypes. However, the active site and zinc binding residues in the different BoNT serotypes were fully conserved (Figures 2.8 and 2.9). As shown in the Figure 2.10, the two amino acid cysteines at 430 and 1235 involved in disulfide bonds are fully conserved in the all different BoNT serotypes whereas the other two amino acid cysteines at 454 and 1280 are variable in some of the BoNT serotypes. Previously, three important amino acids for the protein activity were identified experimentally including amino acids Glu262, Phe266 and Tyr 366 (Figure 2.11). Mutagenesis of these amino acids decreased the protein activity [15].

Table 2.5 Summary of information obtained from multiple sequence alignment of the A-G BoNT serotypes.

Single fully conserved residue	1, 8, 9, 10,12, 13, 19, 35, 37, 39, 44, 47, 49, 86, 90, 95, 97, 99, 100, 106, 109, 118, 121, 194, 203, 229, 230, 234, 236, 237, 238, 240,243, 245, 246, 247, 276, 280, 281, 282, 289, 307, 311, 330, 336, 337, 345, 354, 361, 367, 379, 382, 401, 406, 408, 420, 421, 426, 447, 476, 479, 499, 522, 550, 573, 574, 577, 589, 591, 595, 596, 615, 623, 626, 633, 634, 643, 645, 646, 649, 655, 656, 658, 660, 661, 662, 672, 675, 680, 684, 686, 690, 694, 699, 703, 718, 720, 722, 725, 729, 733, 740, 746, 747, 755, 759, 762, 765, 770, 774, 777, 782, 798, 805, 808, 809, 812, 817, 820, 821, 833, 837, 841, 844, 848, 855, 875, 876, 881, 887, 898, 904, 915, 917, 919, 929, 967, 970, 971, 989, 990, 991, 992, 993, 1001, 1002, 1003, 1004, 1015, 1017, 1021, 1031, 1033, 1042, 1044, 1046, 1050, 1051, 1052, 1053, 1054, 1062, 1063, 1064, 1065, 1067, 1073, 1085, 1087, 1104, 1107, 1110, 1111, 1120, 1130, 1133, 1134, 1137, 1143, 1144, 1147, 1207, 1272, 1332, 1335, 1340, 1341, 1343,
Weakly conserved	2, 11, 17, 42, 57, 63, 64, 102, 122, 125, 164, 210, 286, 287, 331, 353, 356, 413, 422, 446, 484, 486, 487, 489, 503, 511, 569, 579, 582, 586, 590, 607, 609, 612, 613, 616, 618, 620, 621, 642, 681, 621, 746, 757, 760, 771, 773, 780, 784, 803, 816, 832, 835, 858, 863, 883,884, 892, 893, 896, 909, 911, 827, 937, 943, 969, 975, 980, 983, 994, 995, 1014, 1058, 1060, 1070, 1082, 1141, 1205, 1212, 1219, 1232, 1249, 1261, 1287, 1309, 1329, 1330, 1333, 1336,
Strongly conserved	6, 14, 15, 22, 36, 43, 45, 46, 48, 60, 82, 83, 85, 87, 89, 91, 96, 98, 110, 113, 119, 130, 139, 159, 160, 161, 168, 170, 193, 199, 209, 211, 226, 232, 262, 277, 279, 296, 304, 326, 339, 349, 350, 357, 358, 366, 374, 376, 383, 396, 397, 405, 407, 409, 414, 439, 472, 474, 481, 492, 493, 497, 517, 518, 519, 545, 549, 572, 576, 588, 593, 601, 605, 606, 622, 630, 631, 637, 547, 548, 550, 551, 553, 554, 557, 663, 669, 671, 683, 685, 691, 693, 696, 701, 705, 711, 715, 717, 724, 728, 731, 735, 736, 741, 742, 744, 748, 759, 764, 767, 775, 778, 781, 785, 790, 793, 794, 799, 801, 806, 810, 813, 818, 819, 822, 823, 825, 830, 836, 845, 846, 851, 860, 866, 870, 874, 878, 884, 891, 903, 906, 914, 921, 923, 925, 938, 939, 941, 948, 949, 957, 958, 959, 963, 965, 972, 973, 976, 996, 999, 1000, 1005, 1007, 1019, 1027, 1029, 1040, 1045, 1047, 1049, 1054, 1056, 1061, 1075, 1076, 1079, 1089, 1099, 1101, 1102,1106, 1115, 1116, 1123, 1129, 1131, 1132, 1135, 1139, 1140, 1145, 1146, 1152, 1154, 1155, 1197, 1208, 1209, 1210, 1230, 1240, 1243, 1260, 1286, 1312, 1338
Conserved regions	8-10/ 12-13/ 236-238/ 245-247/ 280-282/ 336-337/ 420-421/ 573-574/ 595-596/ 633-634/ 645-646/ 655-656/ 660-662/ 808-809/ 820-821/ 875-876/ 970-971/ 989-993/ 1001-1004/ 1050-1054/ 1062-1067/ 1110-1111/ 1133-1134/

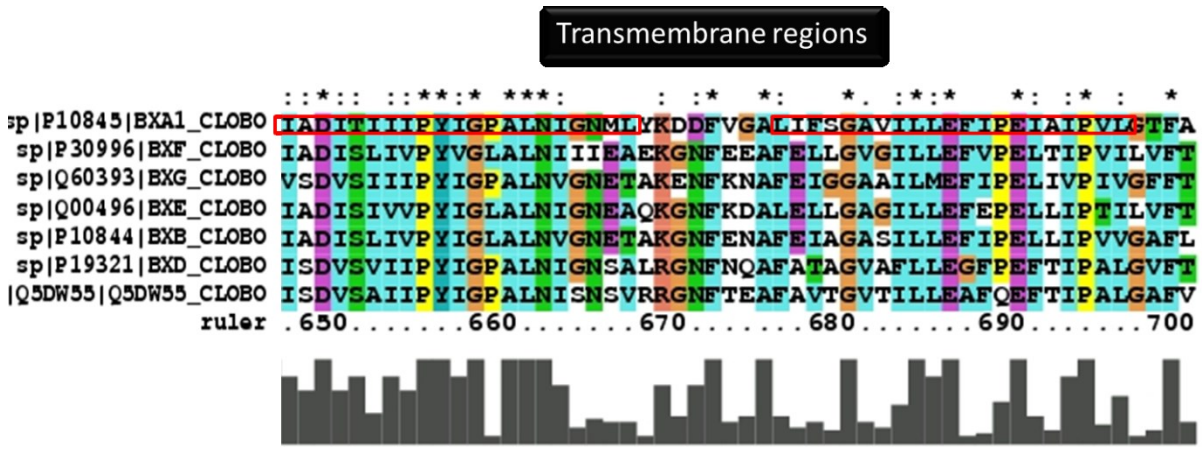


Figure 2.7 Multiple sequence alignment of the different BoNT serotypes. The transmemberane regions of BoNTA are shown in red boxes.

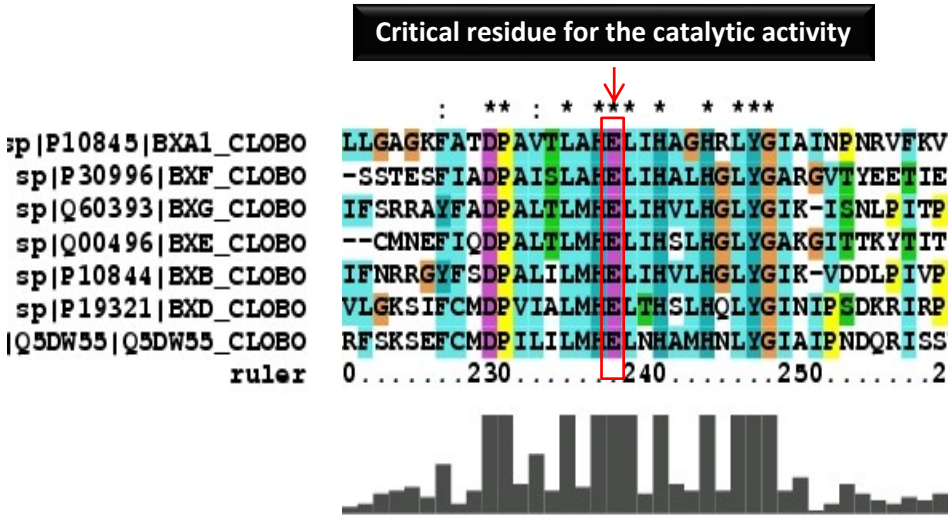


Figure 2.8 Multiple sequence alignment of the A-G BoNT serotypes. The fully conserved EXXX in the active site in the different BoNT serotypes is shown by an arrow. BoNT serotypes are as follow: Q60393: BoNTG, P10844: BoNTB, P10845: BoNTA, P30996: BoNTF, Q00496: BoNTE, P19321: BoNTD and Q5DW55: BoNTC.

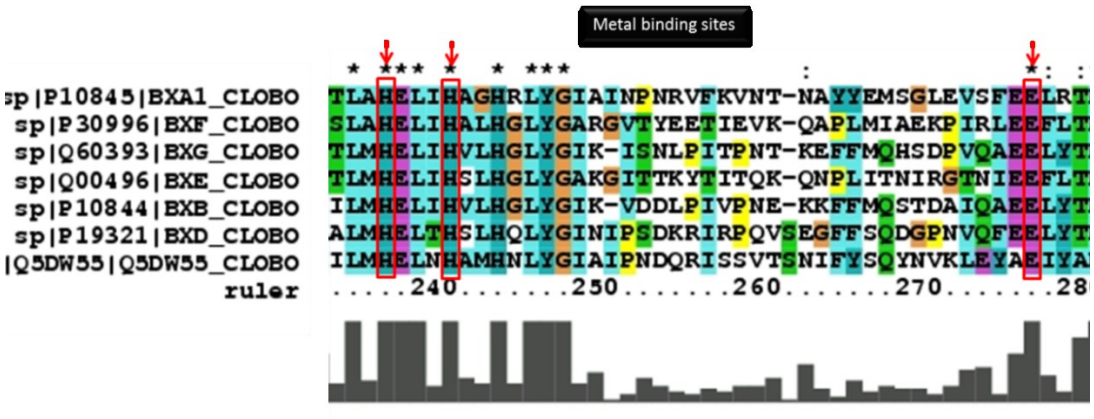


Figure 2.9 Multiple sequence alignment of the A-G BoNT serotypes. The fully conserved zinc coordinating amino acids of the different BoNT serotypes are shown by the arrows. BoNT serotypes are as follow: Q60393: BoNTG, P10844: BoNTB, P10845: BoNTA, P30996: BoNTF, Q00496: BoNTE, P19321: BoNTD and Q5DW55: BoNTC.

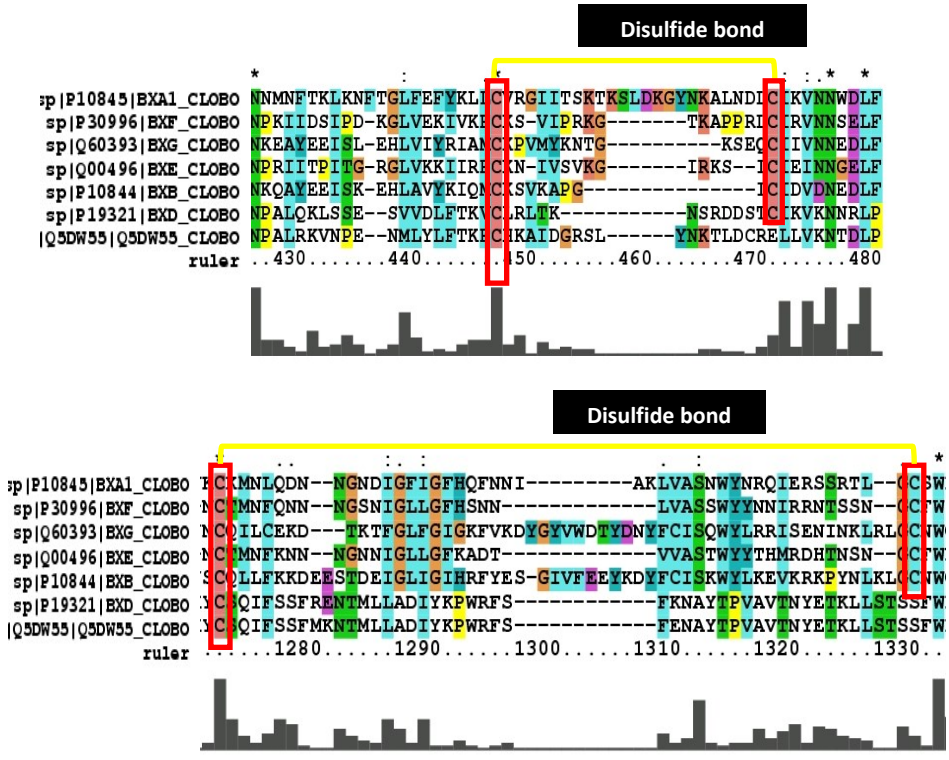


Figure 2.10 Multiple sequence alignment of the A-G BoNT serotypes. The amino acids involved in the disulfide bonds of the different BoNT serotypes are shown in the red boxes. BoNT serotypes are as follow: Q60393: BoNTG, P10844: BoNTB, P10845: BoNTA, P30996: BoNTF, Q00496: BoNTE, P19321: BoNTD and Q5DW55: BoNTC.

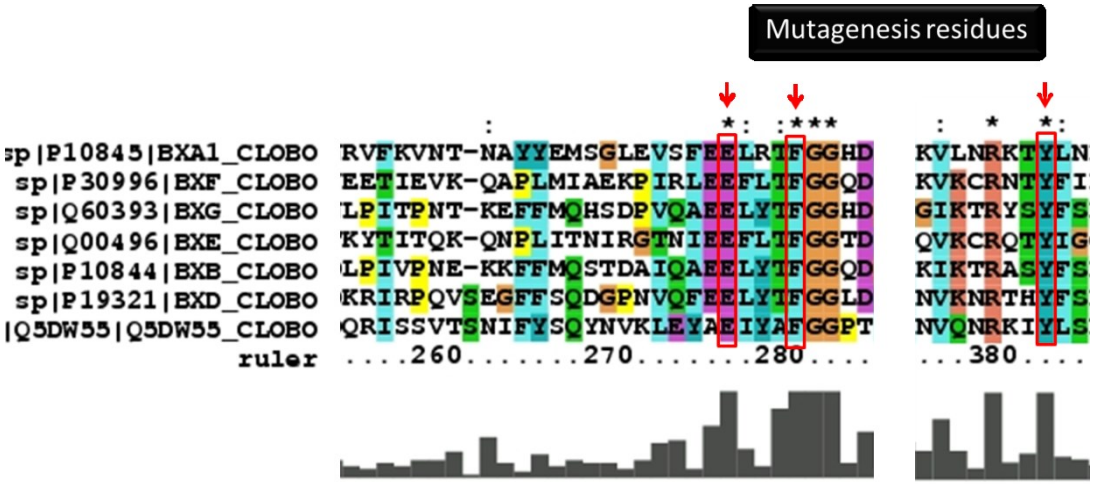


Figure 2.11 The mutagenesis amino acids in the different BoNT serotypes are shown in the red boxes. These amino acids are important due to mutagenesis of them decrease (Phe 266 and Tyr 366) or prohibit (Glu 262) BoNTs proteolytic activity. BoNT serotypes are as follow: Q60393: BoNTG, P10844: BoNTB, P10845: BoNTA, P30996: BoNTF, Q00496: BoNTE, P19321: BoNTD and Q5DW55: BoNTC.

2.3.2.4 Identification of conserved motifs

The BoNT serotypes were characterized by the zinc protease motif pattern using PROSITE. The results showed that BoNT serotypes belong to peptidase M27 family [4]. The pattern in each of the BoNT serotypes sequence is shown in the Table 2.6. Based on the Pfam database, the BoNT serotypes consist of four domains including peptidase M27, transmembrane, N-terminal and C-terminal domains. The arrangement of the Pfam domains found in BoNTA is shown in the Figure 2.12A and B [12]. Also, the information about the domains including the starting and ending amino acids as well as the length of the each domain in all of the BoNT serotypes are shown in the Table 2.7.

To determine the conserved motifs of the different BoNT serotypes, the PRINTS database was searched. The identifier for all BoNT serotypes was called BONTOXILYSIN. BoNT serotypes (A-G) are consist of six motifs and the location of each motif is shown in Figure 2.13 [9].

The length of motif 1 and 2, 3, and 6 were 17 amino acids. The length of motif 4, and 5 were 23 and 13, respectively. Figure 2.14 represents the location of domains and motifs in the 3D structure of BoNTA.

Table 2.6 The pattern signature of each different BoNT serotypes is shown in box.

Accession #	Pattern
P10845 Botulinum neurotoxin type A	190 200 210 220 230 240 GSTQYIRFSP DFTFGFEESL EVDTNPLLGA GKFATDPAV[T LAHELIHAG]H RLYGIAINPN
P30996 Botulinum neurotoxin type F	190 200 210 220 230 240 DPSNYGFGSI NIVTFSPEYE YTFNDISGGH NSSTESFIAD PAI[SLAHELI HAL]HGLYGAR
Q60393 Botulinum neurotoxin type G	190 200 210 220 230 240 SPISEGFGAR MMIRFCPSCL NVFNNVQENK DTSIFSRRAY FADPAL[TLMH ELIHVL]HGLY
Q00496 Botulinum neurotoxin type E	190 200 210 220 230 240 IAIVTFSPEY SFRFNDNCMN EFIQDPAL[TL MHELIHSL]HG LYGAKGITTK YTITQKQNPL
P10844 Botulinum neurotoxin type B	190 200 210 220 230 240 FASREGFGGI MQMKFCPEYV SVFNNVQENK GASIFNRRGY FSDPAL[ILMH ELIHVL]HGLY
P19321 Botulinum neurotoxin type D	190 200 210 220 230 240 PSFEGFGTLS ILKVAPEFLL TFSDVTSNQS SAVLGKSIFC MDPVI[ALMHE LTHSL]HQLYG
Q5DW55 Type C neurotoxin	190 200 210 220 230 240 AAQEGFGALS IISISPRFML TYSNATNNVG EGRFSKSEFC MDPI[LILMHE LNHAM]HNLYG -

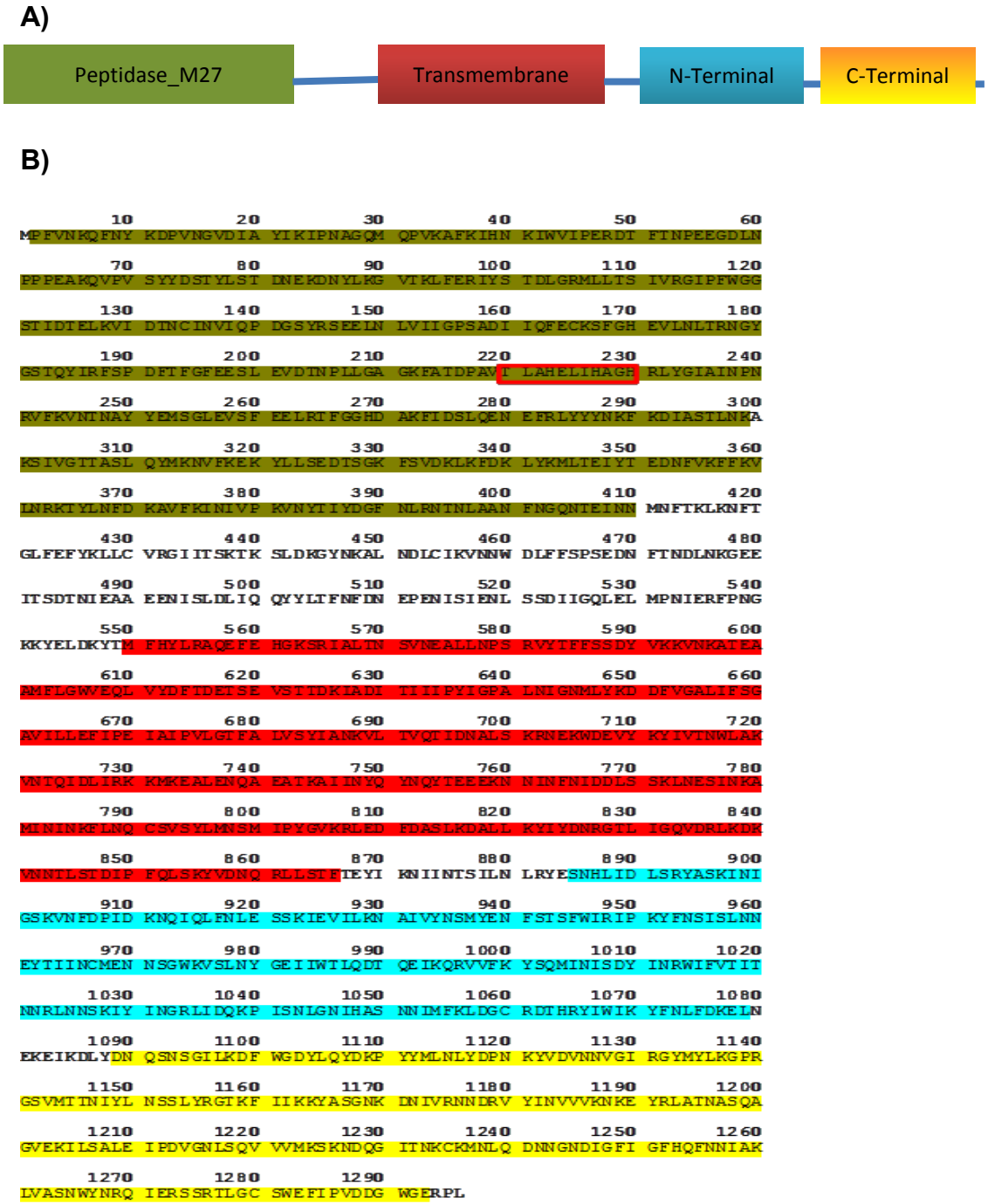


Figure 2.12 BoNTA was chosen to represent BoNT serotypes. BoNTA consists of four domains. The arrangement of the Pfam domains in (A) an overall view of BoNTA sequence and (B) within the protein sequence. Each color represents one specific domain. Catalytic (peptidase_M27), transmembrane, C-terminal and N-terminal domains are highlighted by green, red, yellow and blue colors, respectively. Neutral zinc metallopeptidase region signature (PATTERN) in BoNT sequence is shown in the red box.

Figure 2.13 BoNTs consist of six motifs which associated with the secondary structure of BoNTs. Motifs are shown as color boxes in the multiple sequence alignment of BoNT serotypes. The light blue box located in the motif 5 represents the zinc endopeptidase pattern. BoNT serotypes are as follow: Q60393: BoNTG, P10844: BoNTB, P10845: BoNTA, P30996: BoNTF, Q00496: BoNTE, P19321: BoNTD and Q5DW55: BoNTC.

Table 2.7 Length of domains in the A-G BoNT serotypes.

Serotypes	Peptidase_M27 domain	Transmembrane domain	N-terminal domain	C-terminal domain
BoNTA	2-410 (408 aa)	549-866 (317 aa)	885-1079 (194 aa)	1088-1293 (205 aa)
BoNTB	3-418 (415 aa)	536-853 (317 aa)	872-1066 (194 aa)	1075-1291 (216 aa)
BoNTC	3-419 (416 aa)	545-861 (316 aa)	880-1071 (191 aa)	1080-1280 (200 aa)
BoNTD	3-419 (416 aa)	541-857 (316 aa)	879-1067 (188 aa)	1076-1276 (200 aa)
BoNTE	2-393 (391 aa)	518-840 (322 aa)	859-1054 (195 aa)	1063-1250 (187 aa)
BoNTF	3-410 (407 aa)	537-858 (321 aa)	877-1074 (197 aa)	1089-1274 (185 aa)
BoNTG	3-417 (414 aa)	541-858 (317 aa)	877-1074 (197 aa)	1083-1297 (214 aa)

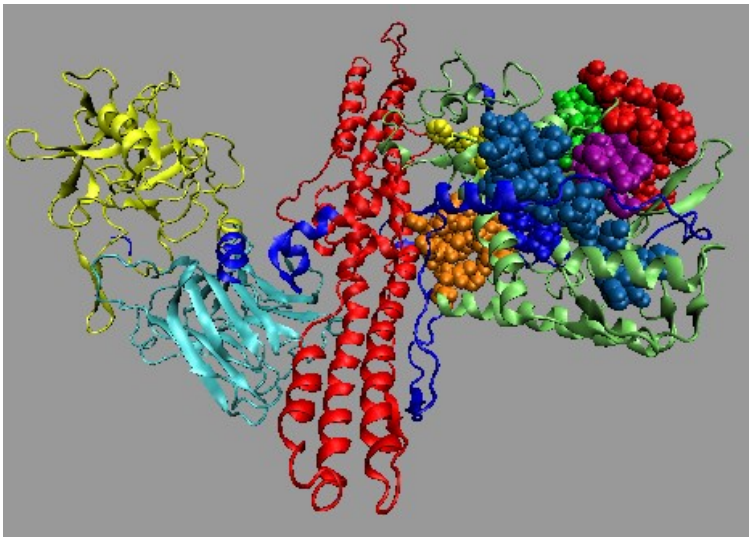


Figure 2.14 The location of domains and motifs is shown in the 3D structure of BoNTA. The green, red, light blue and yellow colors are representing Peptidase-M27, transmembrane, N-terminal, and C-terminal domains, respectively. In addition, red, purple, green, light blue, yellow, and orange colors in VDW drawing represent the motifs (1-6), respectively. Also the dark blue color in VDW drawing represents pattern sequence.

2.3.2.5 Visualization of variable residues in BoNT serotypes

The 3D structure of BoNTA (as a model) was visualized using the 3BTA pdb file in VMD. The variable regions between different serotypes of BoNT were determined using multiple sequence alignment. The variable residues were only visualized in the catalytic domain (Table 2.5) of BoNT serotypes (A-G) due to the high number of variable residues in the entire structure

(Figure 2.15). Additionally based on the BoNT serotypes alignment, all the variable residues in serotype A comparing to other serotypes were determined and listed in (Table 2.8). These residues can be key residues for selective detection of BoNTA serotype from other serotypes. In other words, if a ligand binds to some of these residues, this ligand can be used as a recognition agent to identify BoNTA from other serotypes of BoNT, because these residues are not available in the other serotypes at these specific locations. To achieve this aim, all the potential binding sites in the 3D structure of BoNTA (2NYY) were investigated computationally using the program YASARA (www.yasara.org). The amino acids located within 7-10 Å from the center of the binding regions identified by the program Pass [22] were determined for each of the 33 potentials binding sites within the BoNTA structure (Table 2.9). This table was used to prioritize binding sites with the higher number of key residues for molecular docking (Table 2.8). The binding regions identified as having selectivity to BoNTA are R16, R23, R8, R30, R13, R3, R19, R21, and R22. The region 16 is the most selective one with 4 key residue within 7 A from the center of the region. The analysis of molecular docking results will thus focus on these regions, in order to identify subtype-specific recognition agents.

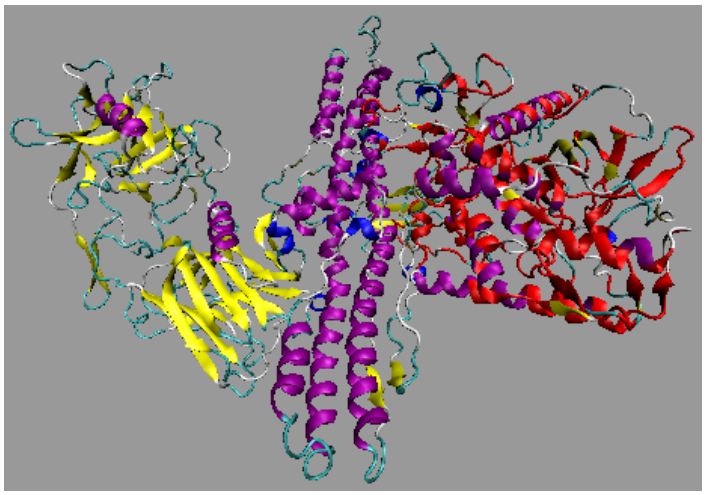


Figure 2.15 Variable residues in the catalytic domain between BoNT serotypes. The red colors represent the variable residues found in the catalytic domain (only) of BoNT serotypes.

Table 2.8 The variable amino acids of BoNTA based on multiple sequence alignment data are listed in this table.

BoNTA variable residue			The amino acid Replaced in other serotypes		Comments	
Number of residue in the alignment	**Name of residue**	**Amino acid group**	**Name of residue**	**Amino acid group**		**Number of Amino acid in the sequence**
5	Asparagine ASN	Polar	Isoleucine	Nonpolar	Except BoNTD replaced by Valine	5
56	Glutamic acid GLU	Polar/-	Proline	Nonpolar		55
77	Serine	polar	Proline	nonpolar		75
101	Tyrosine TYR	Polar	Asparagine	Polar	Except BoNTB replaced by Lysine	99
126	Isoleucine ILE	Nonpolar	Proline	Nonpolar		123
130	Leucine LEU	nonpolar	Phenylalanine	nonpolar		127
166	Alanine ALA	Nonpolar	Proline	nonpolar	Except BoNTC replaced by Glutamic acid	158
193	Tyrosine TYR	Polar uncharged	Phenyl alanine	Nonpolar		180
242	Glycine GLY	Nonpolar	Leucine	Nonpolar	Except BoNTC replaced by Methionine (nonpolar)	229
278	Arginine ARG	Polar/ +	Tyrosine	Polar uncharged	Except BoNTF and E replaced by Leucine (Nonpolar)	264
301	Tyrosine TYR	Polar uncharged	Leucine	Nonpolar		287
310	Threonine THR	Polar uncharged	Arginine	Polar/ +	Except BoNTE replaced by Lysine (Polar/ +)	296
329	Methionine MET	nonpolar	Tyrosine	Polar uncharged		313
365	Tyrosine TYR	Polar uncharged	Phenyl alanine	Nonpolar	Except BoNTD replaced by Methionine (nonpolar)	349
378	Leucine LEU	nonpolar	Lysine	Polar/+	Except BoNTC replaced by Glutamine polar/_	361
396	Valine VAL	Nonpolar	Leucine	Nonpolar		379
397	Proline PRO	Nonpolar	Aspartic acid	Polar/-	Except BoNTE replaced by Asparagine (Polar uncharged)	380
400	Asparagine ASN	Polar uncharged	Leucine	Nonpolar	Except BoNTC replaced by Valine (Nonpolar)	383
427	Asparagine ASN		Proline		Except BoNTG &B replaced by Lysine	410
483	Proline PRO	Nonpolar	Alanine	Nonpolar	Except BoNTC replaced by Glycine (Nonpolar)	466
487	Asparagine ASN	Polar uncharged	Serine	Polar uncharged	Except BoNTC replaced by Aspartic acid	470
518	Glutamine GLN	Polar uncharged	Aspartic acid	Polar/-	Except BoNTB replaced by Asparagine	500
521	Tyrosine TYR	Polar uncharged	Isoleucine	Nonpolar		503
580	Glutamic acid GLU	Polar/-	Proline	nonpolar	Except BoNTC &D replaced by Serine polar	560
585	Arginine ARG	Polar/ +	Asparagine	Polar uncharged	Except BoNTB &C replaced by	565
601	Arginine ARG	Polar/ +	Lysine	Polar/ +		581
618	Threonine THR	Polar uncharged	Valine	Nonpolar		598
622	Methionine MET	nonpolar	Leucine	Nonpolar		602

Continued on next page

Table 2.8 continued

BoNTA variable residue			The amino acid Replaced in other serotypes		Comments	
Number of residue in the alignment	**Name of residue**	**Amino acid group**	**Name of residue**	**Number of residue in the alignment**	**Name of residue**	**Amino acid group**
641	Valine VAL	Nonpolar	Lysine	Polar/ +	Except BoNTB replaced by Serine	621
651	Threonine THR	Polar uncharged	Serine	Polar uncharged		631
666	Methionine MET	Nonpolar	Glutamic acid	Polar/-	Except BoNTC &D replaced by Serine	646
670	Aspartic acid ASP	Polar/-	Glycine	Nonpolar	Except BoNTG replaced by Glutamic acid	650
671	Aspartic acid ASP	Polar/-	Arginine	Polar/ +		651
676	Leucine LEU	Nonpolar	Phenyl alanine	Nonpolar		656
677	Isoleucine	Nonpolar	Glutamic acid	Polar/-	Except BoNTC &D replaced by Alanine	657
715	Valine VAL	Nonpolar	Isoleucine	Nonpolar		692
716	Glutamine GLN	Polar/-	Lysine	Polar/ +	Except BoNTG replaced by Methionine (nonpolar)	693
742	Alanine ALA	Nonpolar	Serine	Polar uncharged	Except BoNTF and E replaced by Threonine (Polar uncharged)	719
748	Isoleucine ILE	Nonpolar	Phenylalanine	Nonpolar		725
749	Aspartic acid ASP	Polar/-	Asparagine	Polar uncharged	Except BoNTB &G replaced by Tyrosine	726
753	Lysine LYS	Polar/ +	Glutamic acid		Except BoNTC &D replaced by Tyrosine	730
756	Lysine LYS	Polar/ +	Tyrosine	Polar uncharged		733
766	Threonine THR	Polar uncharged	Isoleucine	Nonpolar	Except BoNTB replaced by Leucine (nonpolar)	743
832	Arginine ARG	Polar/ +	Lysine	Polar/ +	Except BoNTC &D replaced by Glutamic acid	807
846	Lysine LYS	Polar/ +	Asparagine	Polar uncharged	Except BoNTG replaced by Glutamic acid	821
854	Threonine THE	Polar uncharged	Isoleucine	Nonpolar	Except BoNTG &B replaced by Tyrosine	829
882	Valine VAL	Nonpolar	Threonine	Polar uncharged		857
886	Arginine ARG	Polar/ +	Isoleucine	Nonpolar	Except BoNTD and C replaced by Leucine (nonpolar)	861
888	Leucine LEU	Nonpolar	Isoleucine	Nonpolar	Except BoNTC &D replaced by Lysine	863
892	Threonine THR	Polar uncharged	Asparagine	Polar uncharged		867
901	Threonine THR		Serine		Except BoNTB &G replaced by Asparagine	876
918	Arginine ARG	Polar/ +	Glycine	Nonpolar		893
932	Aspartic acid ASP	Polar/-	Asparagine	polar	Except BoNTF &E replaced by Tyrosine/polar	907
939	Isoleucine ILE	Nonpolar	Phenylalanine	Nonpolar		914
944	Leucine LEU	Nonpolar	Serine	Polar uncharged	Except BoNTE replaced by Aspartic acid	919
954	Lysine LYS	Polar/ +	Asparagine	Polar uncharged	Except BoNTG replaced by Serine	929
968	Threonine THR	Polar uncharged	Isoleucine	Nonpolar	Except BoNTC &D&B replaced by Valine	943

Continued on next page

Table 2.8 continued

BoNTA variable residue			The amino acid Replaced in other serotypes			Comments	
Number of residue in the alignment	Name of residue	Amino acid group	Name of residue	Number of residue in the alignment	Name of residue		Amino acid group
1135	Aspartic acid ASP	Polar/-	Asparagine	Polar uncharged			1103
1155	Valine VAL	Nonpolar	Isoleucine	Nonpolar			1123
1169	Lysine LYS	Polar/ +	Arginine	Polar/ +		Except BoNTC &D replaced by aspartic acid	1137
1231	Alanine ALA	nonpolar	Tyrosine	polar			1194
1276	Leucine LEU	Nonpolar	Phenylalanine	Nonpolar		Except BoNTG replaced by Cysteine	1239
1288	Phenylalanine PHE	Nonpolar	Leucine	Nonpolar		Except BoNTC &D replaced by Aspartic acid	1249

Table 2.9 The amino acids located within 7-10 Å from center of each 33 potential binding sites are summarized. The highlighted ones are important to differentiate BoNTA from the other serotypes. The highlighted residues are non-conserved amino acids. The yellow highlighted ones are still remained in the same group of amino acid while the blue highlighted ones are from different group of amino acid.

Region	The residue within 7 Å of the regions						The residue within 8 Å of the regions					The residue within 10 Å of the regions							
H1	PHE 163	GLU 164	CYS 165	HIS 223	GLU 224	HIS 227	GLN 162	LYS 166	PRO 239	GLU 257	THR 265	SER 167	GLN 184	TYR 185	ILE 186	ARG 187	ALA 228	His 230	ALA 236
	ARG 231	SER 259	GLU 261	GLU 262	MET 531							ILE 237	ASN 238	VAL 258	PHE 260	LEU 263	MET 529	PRO 532	
H2	GLU 279	ARG 283	ASP 339	LYS 340	LYS 343	GLU 347	ASN 280	LYS 335	LEU 336	TYR 342	TYR 502	LYS 337	PHE 338	LEU 341	MET 344	THR 346	ASN 486	ILE 487	GLU 488
	GLU 488	ALA 490	GLU 491	GLU 492								ALA 489	ALA 490	GLU 493	ILE 494	LEU 498	LYS 71		
H3	PHE 658	SER 659	ILE 889	ASP 890	ARG 893	TYR 894	LEU 888	LEU 891	SER 892			VAL 662	SER 799	TYR 803	LEU 879	LEU 881	HIS 887	ILE 889	ASP 890
	ALA 895	SER 896	LYS 897	ILE 898	LYS 929							VAL 926							
H4	TRP 946	ARG 1013	TRP 1014	TYR 1088	LYS 1098		ASP 1009	ILE 1015	ASN 1032	SER 1092		ILE 947	ARG 948	SER 1008	TYR 1010	ASN 1012	PHE 1016	LYS 1070	GLN 1091
	ASP 1099	PHE 1100	TRP 1101	GLY 1102	ASP 1103		LYS 1236					ASN 1093	SER 1094	ILE 1096	TYR 1104	LEU 1105	LEU 1116	THR 1232	ASN 1233
	LYS 1234	GLU 1283										TRP 1282	PHE 1284						
H5	SER 276	ASN 280	GLU 488	GLU 491	GLU 704		GLU 279	ARG 283	ALA 490	LYS 701	ASN 703	ASP 275	LEU 277	ASP 474	LEU 475	ASN 486	ILE 487	ALA 489	GLU 491
	LYS 705	ASP 707	GLU 708	LYS 711	VAL 68		VAL 709	TYR 856				GLU 492	SER 700	ARG 702	GLU 704	TRP 706	TYR 710	TYR 712	LYS 855

Continued on next page

Table 2.9 continued

Region	The residue within 7 Å of the regions						The residue within 8 Å of the regions					The residue within 10 Å of the regions							
H6	ILE 161	GLN 162	PHE 163	PHE 194	THR 220		VAL 70	GLU 351				LYS 66	PRO 69	ASP 159	ILE 160	GLU 164	CYS 165	ARG 187	PHE 188
	HIS 223	GLU 224	GLU 262	ARG 363	ASP 370							THR 193	THR 215	VAL 219	LEU 221	HIS 227	PHE 266	TYR 366	PHE 369
H7	TYR 710	MET 797	ILE 801	VAL 805	ASP 848		TRP 706	PHE 866	TYR 894	LYS 1077		LEU 796	PRO 802	LEU 808	GLU 809	ILE 849	GLN 852	TYR 869	ASN 875
	PRO 850	PHE 851	ILE 870	THR 876	ASP 1076							ILE 878	TYR 934						
H8	ASP 907	ILE 909	GLN 913	ARG 948	TRP 1068		PRO 908	GLN 915	TRP 1014	PRO 1286		ASN 905	PRO 908	ASN 1012	TYR 1066	LYS 1070	TYR 1104	TYR 1111	VAL 1287
	ASP 1099	TRP 1101	ASP 1103	GLN 1106	LYS 1109	ASP 1288						ASP 1288	ASP 1289						
H9	PHE 3	VAL 4	ASN 5	LYS 6	PHE 8		VAL 17	ILE 42				PRO 2	GLN 7	ASN 15	ASP 18	ILE 19	PHE 36	LYS 41	TRP 43
	ASP 18	LYS 37	ILE 38	HIS 39	ASN 40							GLU 147	GLU 511	ASN 514					
H10	PRO 69	VAL 70	GLU 197	PHE 369	ASP 370	LYS 371	LYS 212	THR 420	ASN 459			VAL 68	SER 71	GLY 195	PHE 196	ASN 368	VAL 373	LYS 415	ASN 418
	ALA 372	ASN 418	GLY 421	LEU 422	ASN 458	TRP 460						PHE 423	GLU 424	PHE 425	VAL 457	LYS 731	GLU 734		
	ASP 461																		
H11	ILE 237	PHE 260	ARG 264	ALA 271	ILE 274		LEU 263	SER 276				HIS 230	ILE 235	GLU 261	LYS 272	PHE 273	GLU 281	TYR 286	LYS 343
	ASP 275	GLN 278	GLU 279	ASN 280	PHE 282							THR 350	GLU 491	ILE 494	LYS 711				
	ARG 283	TYR 342	THR 346	GLU 347															
H12	TYR 1117	GLU 1203	PHE 1252	HIS 1253	GLN 1254		GLY 1251	VAL 1262	CYS 1280			ASN 1115	ASP 1118	LEU 1136	ASN 1256	ALA 1259	LYS 1260	SER 1264	GLN 1270
	PHE 1255	TRP 1266	TYR 1267	LEU 1278	GLY 1279							ARG 1276	THR 1277	SER 1281					
H13	PHE 559	GLU 560	GLY 562	LYS 563	SER 564		HIS 561	GLU 741	ALA 742	ILE 747	GLN 750	LEU 568	TYR 751	GLN 753					
	ARG 565	ILE 566	ALA 567	LYS 744	ALA 745														
	ILE 746	ASN 748	TYR 749	ASN 752															
H14	LYS 66	GLN 67	VAL 68	GLU 257	ILE 534		ALA 65	GLN 162	GLY 255	LEU 256	PRO 532	GLU 64	VAL 258	PHE 425	TRP 460	ASN 533	PRO 538		
	GLU 535	ARG 536	PHE 537																
H15	LEU 208	GLY 209	ALA 210	ASN 405	GLU 407	ILE 408	LEU 207	GLY 211	ASN 402	ASP 768		PRO 206	LYS 212	PHE 213	THR 406	ASN 409	GLU 741	LYS 744	
	ASP 767	SER 770	SER 771	LYS 772	ASN 774	GLU 775						ILE 766	LEU 769	LEU 773	SER 776	ILE 777	ASN 778		

Continued on next page

Table of 2.9, continued

Region	The residue within 7Å of the regions						The residue within 8Å of the regions					The residue within 10Å of the regions							
H16	PHE 117	TRP 118	GLY 119	GLY 120	SER 121		LYS 128	LEU 173	ASP 292	ALA 300		VAL 129	ILE 130	HIS 170	ILE 293	SER 295			
	THR 122	ILE 123	ASP 124	THR 125	GLU 126							LEU 297	ASN 298	LYS 301	SER 302				
	LEU 127	TYR 180	THR 296	LYS 299															
H17	ASN 246	TYR 251	GLU 252	SER 465	PRO 466							VAL 245	THR 247	ALA 249	PHE 463	PHE 471	LEU 665	ILE 671	
	SER 467	GLU 468	PHE 667	ILE 668	PRO 669							ALA 719	VAL 721	GLN 724	TYR 795				
	GLU 670	LYS 720																	
H18	VAL 242	PHE 243	LYS 244	VAL 245	ASN 246	THR 247	MET 253	SER 254				GLN 67	ASN 240	ARG 241	ASN 248	SER 259	TYR 366	PHE 464	
	GLY 255	LEU 256	GLU 257	VAL 258	ILE 534	PHE 537						SER 465	PRO 466	ASN 470	GLU 535	ARG 536	PRO 538		
H19	LEU 200	GLU 201	ASP 203	THR 204	LYS 364		GLU 198	VAL 202	LEU 361	ASP 858	LEU 862	SER 199	ASN 205	ASP 352	VAL 355	LYS 356	VAL 360	ASN 362	
	ASN 859	GLN 860	ARG 861									ARG 363	LYS 364	PHE 401	ASP 726	ARG 729	ASN 785	ASN 789	
												ASN 859	LEU 863	SER 864					
H20	PHE 358	LYS 359	VAL 360	ASP 388	GLY 389		ARG 97	TYR 387	ASN 400	GLN 404		ASN 205	ASP 216	VAL 355	PHE 357	LEU 361	THR 385		
	PHE 390	ASN 391	LEU 392	ARG 393	LEU 397							PHE 390	ASN 394	ASN 402	GLY 403				
	ALA 398	ALA 399	PHE 401																
H21	GLY 267	GLY 268	HIS 269	ALA 271	LYS 272	THR 346	ARG 264	PHE 266	LYS 343	GLU 492	TYR 856	LEU 263	THR 265	ASP 270	GLU 279	MET 344	TYR 349	GLU 351	
	GLU 347	ILE 348	THR 350	ASP 352	ASN 353							PHE 354	LYS 364	GLU 491	LEU 496	LEU 498	ILE 499	LYS 855	
	LYS 356	ASN 493	ILE 494	SER 495	ASP 858							TYR 856	VAL 857	ASN 859					
H22	VAL 4	LYS 6	GLN 7	PHE 8	LEU 88		VAL 91	LEU 94				PRO 2	PHE 3	ASN 5	ASN 9	ILE 38	ASP 85	ASN 86	TYR 87
	LYS 89	GLY 90	THR 92	LYS 93	GLU 96							PHE 95	ARG 97	ILE 378	PRO 380				
	VAL 379	LYS 381	ILE 386																
H23	GLU 198	LEU 200	HIS 269	ARG 363	LYS 364		GLU 197	SER 199	ASP 270	ASN 362	TYR 366	PHE 196	GLU 201	PHE 266	PHE 273	ASP 352	LEU 361	PHE 369	
	THR 365	LEU 367	ASN 368	ASN 722	ASP 726		THR 723	ASN 859				LEU 718	ASN 722	GLN 724	ILE 725	ILE 728	ASN 785	ASN 789	
	LEU 727	ARG 729	LYS 730	ARG 861								LEU 862							

Continued on next page

Table of 2.9, continued

Region	The residue within 7Å of the regions						The residue within 8Å of the regions					The residue within 10Å of the regions						
H24	ASP 203	ILE 782	ASN 785	LYS 786	ASN 789		THR 204	ASN 783	PHE 787	GLN 790	GLN 860	LEU 200	GLU 201	VAL 202	ILE 632	ARG 729	MET 781	ILE 784
	ARG 861	SER 864	THR 865				LEU 862					LEU 788	ASN 859	LEU 863	PHE 866	GLU 868		
H25	ARG 177	ALA 236	ILE 237	ASN 238	ARG 241		ASN 178	ILE 235	PRO 239	LYS 289		ASN 174	THR 176	GLY 179	GLY 234	ASN 240	PHE 260	GLU 261
	GLU 281	PHE 282	TYR 285	TYR 286								ARG 264	GLN 278	ARG 283	LEU 284	TYR 342		
H26	TYR 1112	VAL 1123	ASP 1124	VAL 1125	GLY 1157	THR 1158	TYR 1122	ARG 1156	VAL 1184			TYR 1111	MET 1134	LYS 1137	TYR 1155	PHE 1160	LYS 1189	TYR 1191
	LYS 1159	VAL 1185	VAL 1186	LYS 1187	ASN 1188							PRO 1295						
H27	VAL 457	ASN 458	ASP 461	PHE 551	LEU 554		LEU 422	LYS 456	VAL 457	ASN 459	TRP 460	LYS 371	PHE 423	PHE 425	TYR 426	LYS 427	LEU 429	ILE 455
	ARG 555	GLU 558	LYS 731	GLU 734			LEU 462	MET 550	HIS 552	ASN 738		ASP 461	THR 549	PHE 551	TYR 553	ALA 556	GLN 557	ARG 581
												LEU 641	LEU 727	LYS 730	LYS 731	ALA 735		
H28	PHE 679	ALA 680	LEU 681	VAL 682	SER 683		LEU 815	LEU 820	ILE 823	ASN 826		GLU 480	ILE 481	THR 678	PHE 679	TYR 684	THR 691	
	ASP 817	ALA 818	LEU 819	LYS 821	TYR 822							THR 694	ILE 695	ALA 698	SER 814	LYS 816	LEU 820	
	ASP 825											TYR 824	ASP 825					
H29	GLY 1129	ILE 1130	ARG 1131	GLY 1132	TYR 1133		ALA 1197	LEU 1209	LEU 1217	LYS 1260		ASN 1126	ASN 1127	MET 1134	SER 1198	ALA 1200	GLU 1203	ALA 1208
	ALA 1194	THR 1195	ASN 1196	GLN 1199	LYS 1204							SER 1218	LEU 1261	VAL 1262	ALA 1263			
	ILE 1205	LEU 1206	SER 1207															
H30	MET 253	THR 365	LEU 367	ASN 368	PHE 369		LEU 256	TYR 366	TRP 460	ASN 722		VAL 245	TYR 251	GLU 252	HIS 269	ASP 270	PHE 273	
	ASP 461	LEU 462	PHE 463	PHE 464	THR 723		ILE 725	ILE 728				LYS 371	ASN 459	SER 465	MET 550	PHE 667	LEU 718	
	GLN 724	ASP 726	LEU 727	LYS 730								ALA 719	LYS 720	VAL 721	THR 723	ARG 729	LYS 731	
H31	ARG 241	VAL 242	PHE 243	LYS 244	PHE 260	ASP 275	ASN 240	GLU 281	PHE 282	SER 467	GLU 468	ILE 237	ASN 238	VAL 245	VAL 258	SER 259	LEU 263	
	GLN 278	ASP 469	ASN 470	PHE 471	THR 472							PHE 273	ILE 274	LEU 277	PRO 466	THR 472		
H32	ARG 97	SER 100	THR 101	PHE 357	PHE 358		ILE 98	TYR 99	ASP 102	GLY 389		GLU 96	LEU 103	GLY 104	ASP 216	ALA 218	PHE 354	
	LYS 359	ILE 386	TYR 387	ASP 388								VAL 355	LYS 356	VAL 360	THR 385	ARG 393		
H33	LEU 322	LEU 336	LYS 337	LYS 340	LEU 341		PHE 338	ASP 339	TYR 342			MET 106	SER 110	ARG 113	TYR 233	GLU 319	LYS 320	
	MET 344	TYR 502	THR 505	PHE 506								TYR 321	LEU 323	ASP 334	LYS 335	LYS 343	LEU 345	
												ILE 348	TYR 349	GLN 501	TYR 503	LEU 504	ASN 507	
	The residue within 3.5Å						The residue within 7Å					The residue within 8Å						
Active site	HIS 223	GLU 224	HIS 227	PHE 163	GLU 164	CYS 165	GLN 162	THR 220	SER 259	GLU 261		ILE 161	LYS 166	ILE 186	ARG 187	PHE 194	VAL 219	
	GLU 262			THR 265	ARG 363		GLU 351					LEU 221	ALA 222	LEU 225	ILE 226	ALA 228	HIS 230	
												ARG 231	GLU 257	VAL 258	LEU 263	ARG 264	PHE 266	
												TYR 366	MET 531					

2.4 CONCLUSION

In this study, the comparison of BoNTA subtypes sequences indicated that the sequence homology in the BoNTA subtypes is relatively high (%83) with fully conserved amino acids involved in zinc binding and disulfide bonds and highly conserved transmembrane regions. BoNTA-G serotypes were compared. The results showed that the similarity between the serotypes is above 50%. Higher residues variability was seen among the serotypes compared to subtypes of BoNTA, as expected. However, amino acids involving in catalytic activity and zinc binding are still fully conserved among the 7 serotypes. In the case of disulfide bonds, two of the four amino acids were fully conserved in all BoNT serotypes whereas the other two were variable in BoNTD and C serotypes. Although all the serotypes share the same pattern ([G S T A L I V N] - { P C H R } - { K N D } - H - E - [L I V M F Y W] - { D E H R K P } - H - { E K P C } - [L I V M F Y W G S P Q]), the same motifs (1-6) and the same domains, the amino acids in those regions showed some variability between the serotypes. In addition, by exploring the multiple sequence alignment of the BoNT serotypes, we identified nine binding regions with potential selectivity for BoNTA: R16, R23, R8, R30, R13, R3, R19, R21, and R22.

2.5 REFERENCES

1. NCBI, boNT/A1 botulinum neurotoxin type A1 [Clostridium botulinum A str. Hall] . Last modified on March 10, 2011. http://www.ncbi.nlm.nih.gov/gene/5398487. **2011**.

2. NCBI, boNT/A1. Last modified on March 10, 2011. http://www.ncbi.nlm.nih.gov/gene/?term=boNT%2FA1. **2011**.

3. Larkin, M. A.; Blackshields, G.; Brown, N. P.; Chenna, R.; McGettigan, P. A.; McWilliam, H.; Valentin, F.; Wallace, I. M.; Wilm, A.; Lopez, R.; Thompson, J. D.; Gibson, T. J.; Higgins, D. G., Clustal W and Clustal X version 2.0. *Bioinformatics* **2007,***23* (21), 2947-8.

4. Prosite, Neutral zinc metallopeptidases, zinc-binding region signature. Last modified on March 10, 2011. http://www.expasy.org/cgi-bin/nicedoc.pl?PS00142. **2011**.

5. Humphrey, W.; Dalke, A.; Schulten, K., VMD: visual molecular dynamics. *Journal of molecular graphics* **1996,***14* (1), 33-8, 27-8.

6. Bank, P. D., CRYSTAL STRUCTURE OF BOTULINUM NEUROTOXIN SEROTYPE A. Last modified on February 24, 2009. http://www.rcsb.org/pdb/explore/explore.do?structureId=3BTA. **2009**.

7. PDBsum, Toxin. 3BTA. Last modified on March 27, 2011. http://www.ebi.ac.uk/pdbsum/. **1999**.

8. UniPort, boNT. Last modified in 2011. http://www.uniprot.org/uniprot/?query=boNT&sort=score. **2011**.

9. Prints, BONTOXILYSIN. Last modified on June 6, 1999. http://www.bioinf.manchester.ac.uk/cgi-bin/dbbrowser/sprint/searchprintss.cgi?display_opts=Prints&category=None&queryform=false&prints_accn=PR00760. **1997**.

10. UniPortKB, Botulinum neurotoxin type A (A5HZZ9). Last modified on February 8, 2011. http://www.uniprot.org/uniprot/A5HZZ9. **2007**.

11. Arndt, J. W.; Jacobson, M. J.; Abola, E. E.; Forsyth, C. M.; Tepp, W. H.; Marks, J. D.; Johnson, E. A.; Stevens, R. C., A structural perspective of the sequence variability within botulinum neurotoxin subtypes A1-A4. *Journal of molecular biology* **2006,***362* (4), 733-42.

12. Pfam, BXA1_CLOBO. Last modified on March 8, 2011. http://pfam.sanger.ac.uk/protein?acc=P10845. **2007**.

13. PDBsum, PROCHECK summary for 3bta. Last modified on March 27, 2011. http://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/pdbsum/GetPage.pl?template=main.html&o=PROCHECK&c=999&pdbcode=3bta. **1999**.

14. PDBsum, WHAT IF Check report. Last modified in 2009. http://www.cmbi.ru.nl/pdbreport/cgi-bin/nonotes?3bta. **2009**.

15. UniPortKB, Botulinum neurotoxin type A (P10845). Last modified on March 8, 2011. http://www.uniprot.org/uniprot/P10845. **2007**.

16. UniPortKB, Botulinum neurotoxin type F (P30996). Last modified on March 8, 2011. http://www.uniprot.org/uniprot/P30996. **1993**.

17. UniPortKB, Botulinum neurotoxin type G (Q60393). Last modified on March 8, 2011. http://www.uniprot.org/uniprot/Q60393. **2007**.

18. UniPortKB, Botulinum neurotoxin type E (Q00496). Last modified on March 8, 2011. http://www.uniprot.org/uniprot/Q00496. **2007**.

19. UniPortKB, Botulinum neurotoxin type B (P10844). Last modified on March 8, 2011. http://www.uniprot.org/uniprot/P10844. **2007**.

20. UniPortKB, Botulinum neurotoxin type D (P19321). Last modified on March 8, 2011. http://www.uniprot.org/uniprot/P19321. **1990**.

21. UniPortKB, Botulinum neurotoxin type C (Q5DW55). Last modified on March 8, 2011. http://www.uniprot.org/uniprot/Q5DW55. **2005**.

22. Brady, G. P., Jr.; Stouten, P. F., Fast prediction and visualization of protein binding pockets with PASS. *Journal of computer-aided molecular design* **2000,***14* (4), 383-401.

3 CHAPTER THREE: PACLITAXEL IS AN INHIBITOR AND ITS BORON DIPYROMETHENE DERIVATIVE IS A FLUORESCENT RECOGNITION AGENT FOR BOTULINUM NEUROTOXIN SUBTYPE A

Saedeh Dadgar,[1] Zack Ramjan,[2] Wely B. Floriano[1,]*

J. Med. Chem., 2013, 56 (7), pp 2791–2803.

[1]Departmentof Chemistry at Lakehead University and Thunder Bay Regional Research Institute, Thunder Bay ON P7B 5E1, Canada

[2]Biological Sciences Department, California State Polytechnic University Pomona, Pomona CA 91768, USA

[*]CORRESPONDING AUTHOR

SaedehDadgar carried out the experimental parts of the research.

3.1 PREFACE

In this study the FRET assay was employed to identify paclitaxel as an inhibitor (IC_{50} of 5.2 μM) for the catalytic domain of BoNTA. The specific binding of paclitaxel to BoNTA LC was confirmed using scintillation proximity assay. Preliminary confirmation of binding of Pac-BDP (a paclitaxel fluorescent derivatives) to BoNTA LC was obtained in a FP assay.

3.2 ABSTRACT

We have successfully identified one new inhibitor and one new fluorescent recognition agent for the *botulinum* neurotoxin subtype A (BoNTA) using the virtual screening protocol Protein Scanning with Virtual Ligand Screening (PSVLS). Hit selection used an in-house developed Holistic Binding scoring method. Selected hits were tested experimentally for inhibitory activity using Fluorescence Resonance Energy Transfer (FRET) assays against the light chain (catalytic domain) of BoNTA. Ligand binding was determined against the light and heavy chain BoNTA complex either through radiolabeled ligand binding assays (non-fluorescent ligands) or fluorescence intensity assays (fluorescent ligands). These experimental assays have confirmed one compound (paclitaxel) to inhibit BoNTA's proteolytic activity experimentally with an IC_{50} of 5.2 μM. A fluorescent derivative was also confirmed to bind to the toxin, and therefore is a suitable candidate for the rational design of new detection agents and for the development of fluorescence-based multi-probe detection assays.

KEYWORDS: *botulinum* neurotoxin subtype A, botulism, *botulinum* toxin inhibitors, detection agents, fluorescent probes, virtual screening, holistic binding, probe discovery.

3.3 INTRODUCTION

Botulinum neurotoxin (BoNT) is considered the most poisonous toxin known (http://www.cdc.gov). It is extremely potent and lethal, and it is also easy to produce, transport, and use for weaponry.[1-2] Inhalation or ingestion of BoNT causes a neurological syndrome (botulism) that can lead to death by paralysis of the breathing muscles within 24 hours.[3] The toxin prevents the release of the transmitter acetylcholine (ACh) in synapses and neuromuscular junctions, thus impeding communication between neurons and muscle cells, which results in muscle paralysis.[2] The toxin prevents release of the transmitter ACh by cleavage of specific soluble N-ethylmaleimide-sensitive attachment protein receptor (SNARE) through the toxin's light chain (a zinc-dependent endoprotease). All seven BoNT subtypes, labeled A to G, prevent ACh release.[2, 4-8] However, the specific SNARE protein and cleavage site targeted vary among the seven serotypes. Serotypes A, C and E cleave synaptosome-associated protein of 25 kDa (SNAP-25), each at a unique peptide bond. Serotype C also cleaves syntaxin. Serotypes B, D, F, and G target synaptobrevin (also known as vesicle-associated membrane protein). All serotypes require substrates with a minimum of 40 amino acids for efficient cleavage.[9] The subtype A toxin is responsible for the highest mortality rate in botulism.[10]*Botulinum* neurotoxins are synthesized as inactive single chain proteins, cleaved, and released as two-chain (a 100 kDa heavy chain and a 50 kDa light chain) complexes. The heavy (HC) and the light chain (LC) are bound together by one disulfide bond (this complex is referred to throughout the text as whole BoNTA).

The mechanism of action and potency of BoNT makes it an effective therapeutic agent. Successful treatment of over 50 conditions involving hyperactivity of nerves communicating to various muscles or glands have been reported in the literature in recent years. [11-17] These

conditions range from incontinence, spasticity, and focal dystonia (sustained contraction of muscles), to pain. BoNT is particularly useful for managing symptoms of Parkinson's Disease (PD), such as tremors, dystonias, sialorrhea (drooling), and constipation.[13-14, 18-21] However, the most notorious use of BoNT is cosmetic to temporarily reduce the appearance of wrinkles. [22] Under brand names such as Botox® (Allergan, Inc.) and Dysport® (MedicisPharmaceuticalCo.) BoNT injections have become the most performed non-surgical cosmetic procedure in the world. [22-25]Although effective, the clinical use of BoNT is not without risk. Recent findings have shown that locally injected BoNT may affect unintended sites, either directly or indirectly,[26-32] and may induce long-term undesirable effects.[33-34]

BoNT's cosmetic, medical and homeland security importance [35] has fueled an intense search for small molecule inhibitors of its protease activity. Several studies [35-43]have reported novel inhibitors for BoNTA which are promising drug candidates, all of them with activities in the micromolar range. Four small molecule inhibitors based on the 4-amino-7-chloroquinoline scaffold of anti-malarial drugs [36] were discovered using structure-based pharmacophore elucidation and optimization. These inhibitors were found to have IC_{50}s from 3.2 to 17 μM against the *botulinum* neurotoxin subtype A light chain (BoNTA LC).Computer-assisted design of BoNTA-specific hydroxamate inhibitors led to a compound that discriminates between BoNTA LC and BoNTB LC.[42] Pharmacophore search was also used to identify BoNTA LC inhibitors from the National Cancer Institute Open Repository (NCIOR).[37]This search led to the identification of compound NSC 240898 (Kd = 4.6 μM), deemed to be a promising lead for the development of novel therapeutics. A similar search was used to identify another inhibitor from the NCIOR possessing a previously unidentified scaffold, diazachrysene.[41] This scaffold was used to refine existing pharmacophore models, leading to the identification of a non-zinc

coordinating inhibitor with Ki = 600 nM.[38, 41]Peptidomimetics successfully resulted in a potent inhibitor (Ki = 41 nM) of BoNTA LC.[43] High-throughput screening was successful [40] in identifying two lead compounds with *in vivo* activity that represent previously unrecognized chemical scaffolds for the development of therapeutic agents to treat BoNT exposure. Out of 12 compounds found to inhibit BoNTA LC experimentally with IC_{50}s in the 1-95 μM range, only 6 caused 35% or more decrease in SNAP-25 cleavage in cell-based assays. Surprisingly, only 2 of the 12 confirmed inhibitors showed *in vivo* activity.[40] High-throughput screening combined with Structure-activity Relationship (SAR) studies were used to develop inhibitors containing the hydroxamate moiety on different scaffolds in an attempt to minimize the promiscuity for other proteases inherent to this metal chelator.[39] Irreversible inhibitors of BoNTA LC have also been reported. [44]

Despite all the efforts, the discovery and development of new inhibitory agents for therapeutic use has had limited success translating from *in vitro* to *in vivo* activity. [40, 45-46]The inhibitory activity of most of these small molecule inhibitors is based on the coordination of the zinc cation (Zn^{+2}) in the active site of the toxin, [37, 40, 42, 46] which is a common feature of zinc-dependent metalloproteases. These compounds tend to lack sufficient affinity because their binding is mostly due to zinc coordination and not anchored to specific atomic interactions to amino acids within the pocket of the protein. The identification of inhibitors with a different mode of binding is, thus, desirable.

In this paper, we report the successful identification of a new inhibitor and a new fluorescent recognition agent for BoNTA using a computer-assisted discovery methodology. These compounds were identified by an in-house developed methodology, Protein Scanning with Virtual Ligand Screening (PSVLS), based on the *HierVLS* protocol for virtual screening,[47-48] and

a new in-house developed scoring function called Holistic Binding (Floriano et al., in preparation). The fluorescent derivative was shown experimentally to bind to BoNTA complex, the form in which the toxin is distributed throughout the body, and is useful for the development of novel detection assays for BoNTA. The inhibitor reported here represents a previously unidentified class of inhibitors and may lead to development of high-specificity BoNTA inhibitors for therapeutic use. The results presented in this study indicate that Holistic Binding is an efficient approach for the identification of recognition agents for BoNT.

3.4 METHODS AND PROCEDURES

3.4.1 Overall strategy.

The PSVLS methodology was used to computationally screen a library of 1,624 chemical compounds. This virtual library was designed to include compounds readily available commercially with a radioactive label (^{3}H or ^{14}C) to facilitate experimental confirmation of lead candidates. Binding scores for each chemical compound (ligand) in the screening library were calculated using the Holistic Binding scoring function, which requires the molecular docking of each ligand into all available binding pockets within the target protein. Ligands were ranked by Holistic Binding scores and five were selected for experimental confirmation of binding. These included one compound predicted to be a strong inhibitor, two compounds predicted to have borderline affinity for the catalytic site of the toxin, one compound known to delay onset of botulism and predicted to have borderline affinity by Holistic Binding, and one compound predicted to be inactive (a “non-binder”). Selected hits were tested for binding to the target protein and for inhibition of catalytic activity. Ligand binding was determined using radiolabeled ligand binding assays, while inhibition of catalytic activity was determined using Fluorescence Resonance Energy Transfer (FRET) assays. Inhibition assays were performed for the

recombinant light chain of BoNTA, which is catalytically active, while binding assays used non-activated BoNTA heavy and light chain complex. Confirmation of binding to whole BoNTA was important because the catalytic site is expected to be partially occluded in the whole BoNTA,[49-51] which may affect efficacy of the inhibitor *in vivo*. Three commercially available fluorescent derivatives of one confirmed hit (paclitaxel) were docked computationally to whole BoNTA, and tested experimentally for binding to whole BoNTA using fluorescence intensity. The objective was to investigate whether any of these derivatives might be useful as a recognition agent in a detection assay.

In addition to in-house developed software for PSVLS and Holistic Binding score calculations, the programs Yasara (www.yasara.org) and MOE (www.chemcomp.com) were used for generation and manipulation of 3D structures, and for structural analysis of docked complexes. All chemical compounds were purchased from Sigma Aldrich (www.sigmaaldrich.com) and Cedarlane Labs (www.cedarlanelabs.com). Toxin was obtained from List Biological Laboratories (www.listlabs.com) and from Metabiologics (www.metabiologics.com). PSVLS calculations were performed on High Performance Computer Clusters running the Linux operating system.

3.4.2 Computational screening

Selection and preparation of the ligand database to be computationally screened. In order to facilitate the experimental validation of the PSVLS/Holistic Binding scoring methodology, we constructed a database of chemical compounds that are readily available for purchasing with a ^{14}C or a ^{3}H radioactive label. Compounds for virtual screening were selected from the Available Chemicals Directory (ACD) (Symyx, Inc.) through DiscoveryGate (www.discoverygate.com). We searched for compounds containing "H^{3}" or "C^{14}". Chemical structures were downloaded in

sdf format (2D) and converted to three-dimensional (3D) structures using MOE (www.chemcomp.com). Ligand structures were stripped of counter-ions. Gasteiger atomic charges were assigned to each ligand in the database. The ligands were subject to geometry optimization followed by energy minimization using the Dreiding force field. [52] The final optimized 3D structures were saved as multiple-ligands in mol2 format and used for virtual screening.

Preparation of the target protein for structure-based virtual screening. The experimentally-determined 3D structure of the target protein, *botulinum* neurotoxin subtype A, was downloaded from the RCSB Protein Data Bank (www.rcsb.com). We used the structure identified by the PDB code 2NYY (2.6 Å resolution) which corresponds to whole BoNTA (heavy and light chains). The structure was checked for missing atoms and/or segments, and overall structural quality using the software packages Yasara (www.yasara.org) and/or MOE (www.chemcomp.com). Ligands, co-factors, and counter ions present in the pdb file were removed, and missing atoms/segments were added to the structure. Missing segments Ala490-Glu492 and Thr623-Lys626 were copied from PDB structure 3BTA (BoNTA at 3.2 Å resolution). The complete structure was assigned CHARMM22 [53] atomic charges and energy-minimized for 2,000 steps of conjugate-gradient minimization using the Dreiding force field.[52]

Protein Scanning with Virtual Ligand Screening (PSVLS)**.** In this in-house developed approach, the 3D structure of a target protein is scanned computationally for potential binding sites using the program *Pass*.[54] This program is one of many methods available for identification of binding pockets in proteins.[55-64]*Pass* is freely distributed, efficient, and its outputs are compatible with the downstream applications in PSVLS, making it a convenient choice for our approach. *Pass* uses a geometric approach to identify voids in protein structures fast and

accurately. The size, shape and burial extend of these voids are used to assess their likelihood to represent binding pockets. A virtual library of chemical compounds is screened against each available binding pocket identified by Pass within the target protein using molecular docking. Virtual screening is performed using the *HierVLS* protocol,[47] with parameters adjusted to produce more accurate results at the expense of computational time. *HierVLS* is a computational protocol for fast screening of large chemical compound libraries.[47] *HierVLS* uses a hierarchical approach: a large number of bound configurations are created in the least computationally expensive step for each ligand. Subsequent steps reduce the number of bound configurations per ligand while increasing the computational time required to process them, thus increasing accuracy. The last step includes the calculation of the binding energy with solvation for the best surviving bound configuration of each ligand. PSVLS generates 3D structures and calculates force field-based binding energies for each ligand bound to each binding region of the target protein. For each binding region, *i*, force field-based binding energies (*FF_BE)* are calculated as the difference between the sum of potential (*PE*) and solvation (*SE*) energies of the bound state minus the unbound state:

(EQ 1) $$FF\_BE_i = (PE + SE)_{BoundComplex} - (PE + SE)_{UnboundProtein} - (PE + SE)_{UnboundLigand}$$

Ligands with favorable (more negative) calculated binding affinities to the target binding site are potentially active compounds.

The 1,624 ligands in the virtual library were computationally screened against each one of the 33 binding regions identified in the BoNTA structure (pdb code 2NYY) using PSVLS. This virtual screening generated 1,624 X 33 = 53,592 bound structures and respective calculated binding energies for analysis. PSVLS calculations were performed through the *Orunmila*

(formerly known as *Cassandra)*[48] graphical user interface on a Linux-based High Performance Computer Cluster. The average time for each ligand docked into all 33 binding regions of the BoNTA structure was 35 hours on one core of a Intel Core 2 Quad core processor, and each run generated >200 GB of data.

Selection of PSVLS Hits. The force-field based scores calculated in PSVLS for each ligand bound to a targeted site within the protein reflects the affinity of the ligand for that particular site. Once in close proximity of each other, however, on-target but off-site binding events may take place. These events are reflected in the binding constants measured experimentally but are not directly considered, to our knowledge, by any of the current methods for calculating binding affinities. In our Holistic Binding scoring approach, the binding scores for each ligand are calculated as a function of the force field-based binding energies calculated by PSVLS at each available binding pocket. The force field-based binding energies are scaled by a propensity factor (second term in equation (2)) that reflects the ratio between the force field-based binding energy for region *i* and the sum of the force field-based binding energies for all *n* regions:

(EQ 1) $$HBS_i = FF\_BE_i \times \frac{FF\_BE_i}{\sum_{j=1}^{n} FF\_BE_j}$$

The result of this approach is a set of ligands ranked by binding affinity for a particular binding site relative to all other potential sites within the target protein.

The results of PSVLS/Holistic Binding scoring were analyzed in order to identify potential inhibitors targeting the catalytic region of BoNTA. Ligands with less than 70% of their molecular surface buried into the binding region within the protein were eliminated from the list, regardless of binding energy. Overly charged ligands and ligands with phosphate groups were excluded from consideration to avoid bias introduced by the limitations of the force field in describing correctly the inter-atomic interactions for these ligands. Remaining ligands were

ranked according to their calculated Holistic Binding scores. Ligands with better (more negative) Holistic Binding scores are more likely to bind experimentally to the target protein.

3.4.3 Experimental confirmation of selected hits

Inhibition Assay. Catalytic activity of BoNTA was determined using fluorescence resonance energy transfer (FRET) in an assay that requires no separation of free from bound ligands. SNAPtide® (FITC/DABCYL) (List Laboratories, Inc), a synthetic quenched fluorescent peptide containing the native cleavage site for *botulinum* toxin type A, was used in competition assays against the ligands selected for testing. Cleavage of the substrate by BoNTA releases the fluorophore (FITC) and full fluorescence is restored. The increase in fluorescence intensity is directly proportional to the amount of cleavage that has occurred and thus allows for accurate measurement of enzymatic activity.

Light chain BoNTA recombinant protein (BoNTA LC) was purchased from List Laboratories, Inc., and reconstituted in a buffer of 20 mM HEPES and 0.1% TWEEN-20 at pH 8.2. Enzymatic activity was determined prior to ligand binding experiments for 10nM of BoNTA LC in a FRET assay using SNAPtide® unquenched calibration peptide (List Laboratoris, Inc.). The correlation coefficient for a linear fitting of concentration versus Relative Fluorescence Units (RFU) was 0.996 for a concentration range from 0.0125 μM to 0.8 μM of substrate.

SNAPtide® substrate was reconstituted following the manufacturer's instructions. A final concentration of 8 μM SNAPtide® per well was used for the competition assays. Ligands were purchased from Sigma-Aldrich, Inc., and from Cedarlane Laboratories. These were: paclitaxel [CAS 33069-62-4] from *Taxusbrevifolia* (≥95% purity) and semi-synthetic (≥97% purity); D-(-)-fructose [CAS 57-48-7] (≥99% purity); desmosine [CAS 10019-68-8] (99.5%±0.2% purity); aminopterin [CAS 54-62-6] (≥96% purity); monensin [CAS 22373-78-0] (90-95% purity).

Ligands were prepared as 1 mM stock solutions and diluted appropriately. DMSO, deionized water, 2.0M aqueous NaOH, or ethanol was used as solvent, according to the solubility profiles of each ligand. Whenever DMSO was used as solvent, its final concentration in the assay was kept at less than 2% of the volume, to prevent inhibition of cleavage by the solvent. Ligands were tested in triplicate for a fixed 10nM concentration of toxin in a volume of 250 µl per well. For the competition assays, substrate was added to the mixture of toxin and ligand. Assays were performed in 96-well black polystyrene microplates (OptiPlate, PerkinElmer, Shelton, CT, USA, http://las.perkinelmer.com/). FRET assays were run at 37°C after 1 hour of incubation. Excitation wavelength was set to 490 nm and emission was set to 523 nm, with a cutoff filter at 495 nm. Fluorescence intensity was recorded 1 hour after mixing using a Synergy 4 microplate reader. The percent of inhibition caused by each ligand at each concentration tested was calculated relative to the fluorescence intensity measured for uninhibited cleavage of SNAPtide®, which was determined in triplicate for the same concentrations of substrate and toxin used in the competition assays.

Dose-response curve. A dose-response curve was determined for paclitaxel using the FRET assay described above. A final concentration of 8 µM SNAPtide® and 10 nM of BoNTA LC in a total volume of 200 µl per well was used for the competition assays. Assays were conducted in triplicate at 37°C. Paclitaxel and toxin were incubated for 30 minutes prior to the addition of SNAPtide®. The DMSO effect on the assay was taken into account using controls with concentrations varied from 0.0008% to 40%. EDTA (ethylenediaminetetraacetic acid) was used as an inhibitor control at varying concentrations (10 mM-10 pM). EDTA is known to have an inhibitory effect on BoNTA LC enzymatic activity.[65-66]

Immobilization of BoNTA to magnetic beads for separation**.** For assays requiring separation of free from bound ligands, biotinylated BoNTA (heavy and light chains) was immobilized onto streptavidin-coated magnetic purification beads (BcMag-streptavidin from Bio-Clone, Inc., www.bioclon.com). Biotinylated whole BoNTA was purchased (Metabiologics, Inc) as 1 μg/ml solution in standard buffer. The beads were washed three times with deionized water prior to toxin immobilization, and the supernatants were discharged. Magnetic beads (2 μl) and whole BoNTA solutions corresponding to 0.2 μg of toxin were incubated for 30 minutes at room temperature at pH 7.4. Loaded (toxin-attached) beads were washed with buffer (25 mM HEPES, 0.3 mM $ZnCl_2$, and 0.2% TWEEN-20) to remove the excess of protein. Protein concentration was determined for the supernatants of the immobilization and washing steps by absorbance at 280 nm using 200 μl aliquots. The amount of streptavidin released from the beads during washing steps was estimated for aliquots without toxin. Loaded beads were re-suspended in buffer prior to use. For separation, microcentrifuge tubes containing suspended beads were placed in a magnetic separator (Bio-Clone, Inc.) until the supernatant became clear and could be aspirated (2-3 minutes). For all assays, non-specific binding to the beads was determined by mixing ligands with beads, without toxin, under the same conditions of the assays being performed. Total binding was determined based on reading of the beads and supernatant fractions.

Ligand Binding Fluorescence Intensity (FI) Assays**.** Fluorescent paclitaxel derivatives were tested for binding against whole BoNTA using fluorescence intensity assays. The toxin was attached to magnetic beads to facilitate separation. The samples containing ligand and toxin were read relative to reference samples of the fluorescent ligands in buffer. Buffer was used as blank. Non-fluorescent paclitaxel was used as negative control. Ligands were added to toxin-loaded

suspended beads and incubated for 1 hour prior to reading. Beads were separated from supernatant using a magnetic separator. Fluorescence intensity was determined for both fractions. Bead fractions were re-suspended in buffer for reading. Fluorescent derivatives of paclitaxel were purchased from Invitrogen, Inc.: Oregon Green 488 paclitaxel [CAS 301844-13-3] (pac-OG); BODIPY®FL paclitaxel[CAS N/A] (pac-FL); BODIPY® 564/570 paclitaxel [CAS N/A] (pac-BDP). Stock solutions for the ligands were prepared according to manufacturer's recommendations. Ligands were tested in triplicate at 10 μM, for a fixed 10 nM concentration of toxin. For these assays, the whole BoNTA toxin (List Laboratories, Inc) was used. Whole BoNTA was reconstituted in a buffer of 20 mM HEPES, 0.3 mM $ZnCl_2$, and 0.2% TWEEN-20 at pH 8.2. BoNTA was not activated with DTT for this assay. The MW of whole BoNTA is 150,000 Daltons. Ligands and toxin were incubated in the same buffer used to reconstitute the toxin for 1 hour at 37˚C prior to reading. The concentration of toxin was determined by absorbance at 280 nm. Excitation and emission wavelengths were set according to the fluorophores to be tested (excitation/emission): BODIPY® FL (505nm/515nm), BODIPY® 564/570 (565nm/571nm), and Oregon Green 488 (496nm/524nm). The BODIPY® fluorophores are boron dipyrromethene derivatives. BODIPY® FL is 4,4-difluoro-5,7-dimethyl-4-bora-3a,4a-diaza-s-indacene-3-propionic acid, whereas BODIPY® 564/570 is 4,4-difluoro-5-styryl-4-bora-3a,4a-diaza-s-indacene-3-propionic acid. Oregon Green® 488 (2',7'-difluorofluorescein) is a fluorescein analog.

Radiolabeled Ligand Binding (RLB) Assays. We used radiolabeled ligands to probe binding of selected ligands to whole (heavy and light chains) BoNTA. Ligands found to inhibit catalytic activity of BoNTA LC do not necessarily bind to whole BoNTA because the catalytic site is believed to be inaccessible in whole BoNTA,[67] which is the form found in the

bloodstream.[68] Since binding to whole BoNTA may contribute positively to *in vivo* activity of the inhibitors, we tested confirmed inhibitors for binding to the heavy and also to the light chain BoNTA complex.

Radiolabeled ligands were purchased from Sigma (St. Louis, MS): D-fructose-UL-^{14}C at 1.0 mCi/ml, with specific activity of 300 mCi/mmol, and purity ≥95%, as determined using HPLC by the manufacturer; and Paclitaxel-(2-benzoyl-ring-UL-^{14}C) at 0.25 mCi/ml, specific activity of 50.7 mCi/mmol, and purity of 99.9% determined using HPLC by the manufacturer. The assay buffer was 25 mM HEPES with 0.2% TWEEN-20 at pH 8.2. Radiolabeled ligands were incubated with whole BoNTA attached to magnetic beads for 1 hour at room temperature. D-(-)-fructose was tested at 10 µM (1.2 uCi) and paclitaxel was tested at 25 µM (0.5 uCi). The final assay volume was 400 µl per vial. The supernatant fraction was separated from the beads fraction using a magnetic separator. Aliquots of 3 ml of EconoSafe scintillation cocktail (Sigma-Aldrich, Inc.) were added to each vial. Counting was performed using both fractions in duplicate samples, on a Beckman LS6500 scintillation counter. Each vial was counted for 10 minutes. This counting time was determined sufficient to reduce the standard deviation to less than 1% of the mean counts per minute. Total and free radiolabeled ligand concentrations were estimated from counting beads and supernatant fractions. Non-specific binding to the beads was determined as described earlier.

Scintillation Proximity Assay (SPA). We used SPA to confirm the dose-dependent binding of paclitaxel to BoNTA LC. Streptavidin-coated Polyvinyltoluene SPA beads were purchased from PerkinElmer, Inc., and reconstituted in 20 mM HEPES buffer at pH 7.4, 0.1% Tween 20, and 0.3 mM ZnCl2. Paclitaxel [2-benzoyl ring-3H] at 1 mCi/ml, with specific activity of 60 mCi/mmol, and purity of 99% determined using HPLC by the manufacturer was purchased from

American Radiolabeled Chemicals, Inc. (St. Louis, MO). BoNA LC (30 µg) was biotinylated using ChromaLink™ Biotin One-Shot Antibody Labeling Kit (Solulink, Inc). At each ligand concentration, Paclitaxel [2-benzoyl ring-3H] and 7nM biotinylated BoNTA LC were incubated for 15 minutes at room temperature before addition of SPA beads at 2mg/ml per well. The assay was run in triplicate at the final volume of 200 µl per well, with readings at 1 hour intervals. Scintillation was counted on a MicroBeta2 plate counter (PerkinElmer, Inc) and corrected for non-specific binding/non-proximity effects.

3.5 RESULTS AND DISCUSSION

3.5.1 PSVLS/holistic binding scores.

Results from PSVLS consist of 3D structures for each ligand bound into each binding cavity of the target protein, and their respective force field-based binding energies. A total of 53,592 protein ligand complexes were generated for the 1,624 screened ligands bound to 33 binding cavities in the BoNTA structure. The 3D structure for BoNTA used in the PSVLS calculations and the location of the centers for the 33 binding cavities identified in this structure are shown in Figure 3.1. Some of these binding regions correspond to known binding sites for substrates (Catalytic Site), and for polysialogangliosides and protein receptors which bind to the Receptor Binding Domain. The other sites are uncharacterized and may or may not function as allosteric sites. They are, however, cavities large enough to accommodate the binding of a small organic molecule. Holistic Binding scores were calculated for each binding site of BoNTA by combining the force field-based energies, corrected for solvation, according to equation EQ 2. At each binding site targeted for docking, these scores represent the binding affinity of a ligand for that particular site relative to all the other sites. In this paper, we report only the analysis of the Holistic Binding scores for the catalytic site of BoNTA, even though ranked lists of ligands were

also obtained for other sites. The Holistic Binding scores for each one of the 1,624 ligands screened bound to the catalytic site of BoNTA are shown in Figure 3.2. The mean Holistic Score for the entire library in the catalytic site is -1.9 kcal/mol. The mean scores for the other 32 regions ranged from -0.9 kcal/mol to -2.8 kcal/mol.

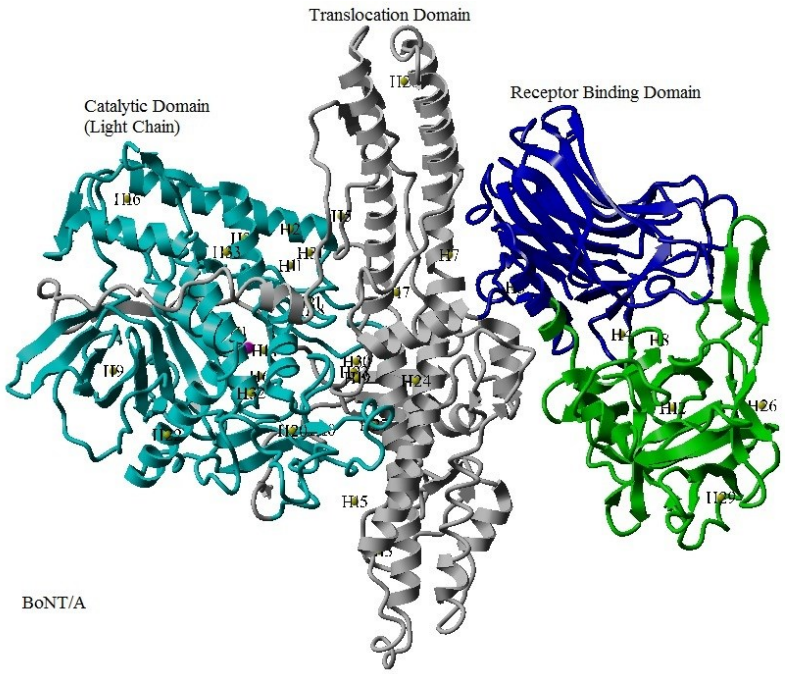


Figure 3.1 Structure of *botulinum* neurotoxin subtype A (BoNTA) (heavy and light chains) and location of the centers for the 33 binding cavities targeted by *PSVLS*. The structure for BoNTA corresponds to pdb code 2NYY. Color scheme: Catalytic Domain (Light chain; residues 2-410) is cyan; Translocation Domain (residues 411-871) is gray; Receptor Binding Domain motif 1 (residues 872-1087) is blue; Receptor Binding Domain motif 2 (residues 1088-1293) is green. Zinc in the Catalytic Site is represented as pink spheres. Yellow spheres represent the centers of binding regions identified and targeted for molecular docking by Protein Scanning with Virtual Ligand Screening (PSVLS). Some of these binding regions correspond to known binding sites for substrates (Catalytic Site), and for polysialogangliosides and protein receptors which bind to the Receptor Binding Domain. This was generated using the molecular modeling software Yasara (www.yasara.org).

Most ligands in a random library are not expected to have specific affinity for the target protein. Using the binding energy definition in equation (EQ 1), these "non-binders" should produce scores equal to or above zero (unbound state being more energetically favorable than

bound state). However, non-specific intermolecular interactions with the protein may produce below zero scores for these "non-binders". In terms of score distribution, this means that the vast majority of ligands in a library will score within certain values while very few ligands, if any, with specificity for the protein target will produce much better scores (more negative, in the case of our scoring function). The threshold score that discriminates binders and non-binders can be estimated from the frequency distribution diagram shown in Figure 3.3. Most ligands (68%) scored near -2 kcal/mol. The vast majority (87.5%) scored ≥-3.0 kcal/mol and should be considered non-binders. If we optimistically consider that 2% or less of the library may be binders, then ligands with scores equal to or lower than -5 kcal/mol should be considered for experimental testing. This value corresponds to the mean of binding energies (show as solid line in Figure 3.2) plus two times the standard deviation for the data set (shown as dashed line in Figure 3.2). This binder (at or below) versus non-binder (above) threshold is region-specific, and must be determined for each binding region of interest. It is also case-specific and may vary with target protein, screening library, scoring function, and/or virtual screening method.

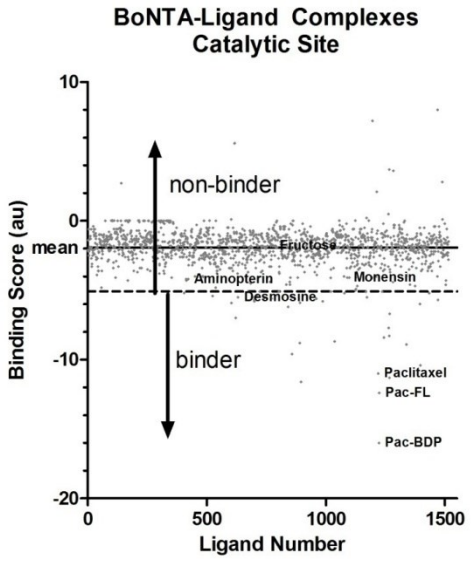


Figure 3.2 Holistic Binding scores at the catalytic site for BoNTA complexes from *PSVLS*. More negative numbers correlate to better binding. Solid line labeled "mean" represents mean value of Holistic Score for the ligand library. Dashed line represents the mean plus two standard deviations (non-binder/binder threshold). Ligands scoring below this threshold may bind experimentally to the target protein.

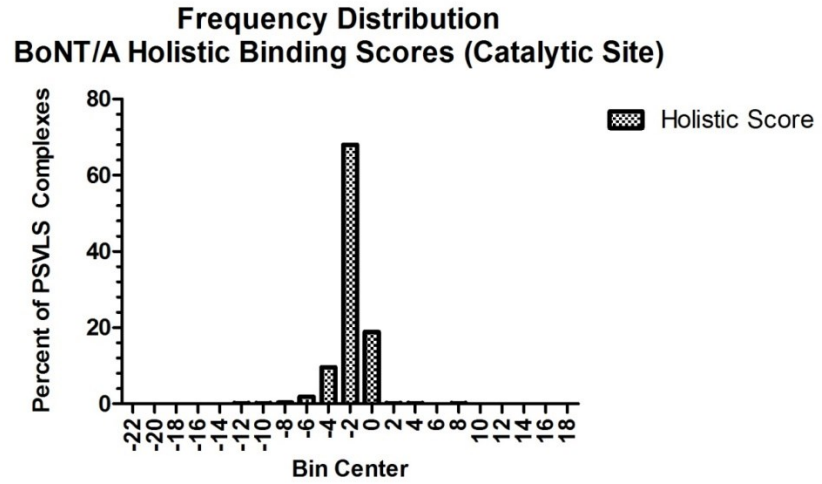


Figure 3.3 Frequency distribution of Holistic Binding scores at the catalytic site. Bars represent the percent of PSVLS complexes that fall into a score bin centered at the plotted value. The Holistic Binding score mean value for this compound library was -1.9 kcal/mol. 87.5% of ligands have scores above -1.0 kcal/mol. 9.6% of the ligands have scores of -4.0±1 kcal/mol. The top 2.8% have scores of -5.0 kcal/mol or better. Biologically active compounds are more likely to be found among the high-ranking (most negative) candidates.

3.5.2 Relative contribution of other regions to the holistic scores

The contribution from FF binding energies in other regions to the Holistic Scores at the catalytic site can be summarized as follows: except for paclitaxel, all of the other ligands selected had 3 to 8 kcal/mol difference between the FF energy at the catalytic site and at the next best non-catalytic site. This translates into less than 1 kcal/mol difference in Holistic Scores. For paclitaxel, the catalytic region has the highest FF binding energy (-90.9 kcal/mol), followed by three non-catalytic regions with FF binding energies around 80 kcal/mol, which are considerably worse than the FF binding energy in the catalytic site. All three non-catalytic regions are at the interface between the catalytic and translocation domains. The differences in FF energies relative to the catalytic site yield Holistic Scores differences from 2.5 to 2.3 kcal/mol. Aminopterin and fructose have better FF binding energies in other regions of the target protein than in the catalytic site, and this trend is preserved in the Holistic Scores.

3.5.3 Selection of PSVLS hits for experimental testing

Five ligands were selected for experimental testing (Figure 3.4). Selection was based on their Holistic Binding scores and ranking at the catalytic site. We chose to focus on the catalytic site because of the existence of a well-established experimental assay to validate the computational predictions about inhibitory activity. After experimental confirmation of the initial 5 ligands, 3 fluorescent derivatives of the best confirmed ligand (paclitaxel) were also docked, scored and tested experimentally for binding. All ligands were re-ranked after inclusion of these fluorescent derivatives to the ranking list. The ligands selected for experimental testing along with the rationale for their selection is presented below. These data are summarized in Table 3.1.

Paclitaxel has the 6th best score in the catalytic site, and the top three (3) scores are its derivatives. The 4th and 5th best ranking ligands were known complexing agents for removal of heavy metal ions and a surfactant. They were expected to have low selectivity for BoNTA for reasons discussed earlier and were, therefore, not selected for experimental testing. The catalytic region has the highest binding affinity to paclitaxel (-11 kcal/mol), followed by three non-catalytic regions (27, 5 and 23) with affinities between -8.5 kcal/mol and -8.7 kcal/mol, which are considerably worse than the binding affinity in the catalytic site. These regions contributed the most for the adjustment of the FF binding energy into the catalytic site holistic score. The score for paclitaxel in the catalytic region is well below the non-binder/binder threshold for the catalytic region (-5 kcal/mol) and, for this reason, we expect paclitaxel to bind strongly and inhibit experimentally BoNTA LC.

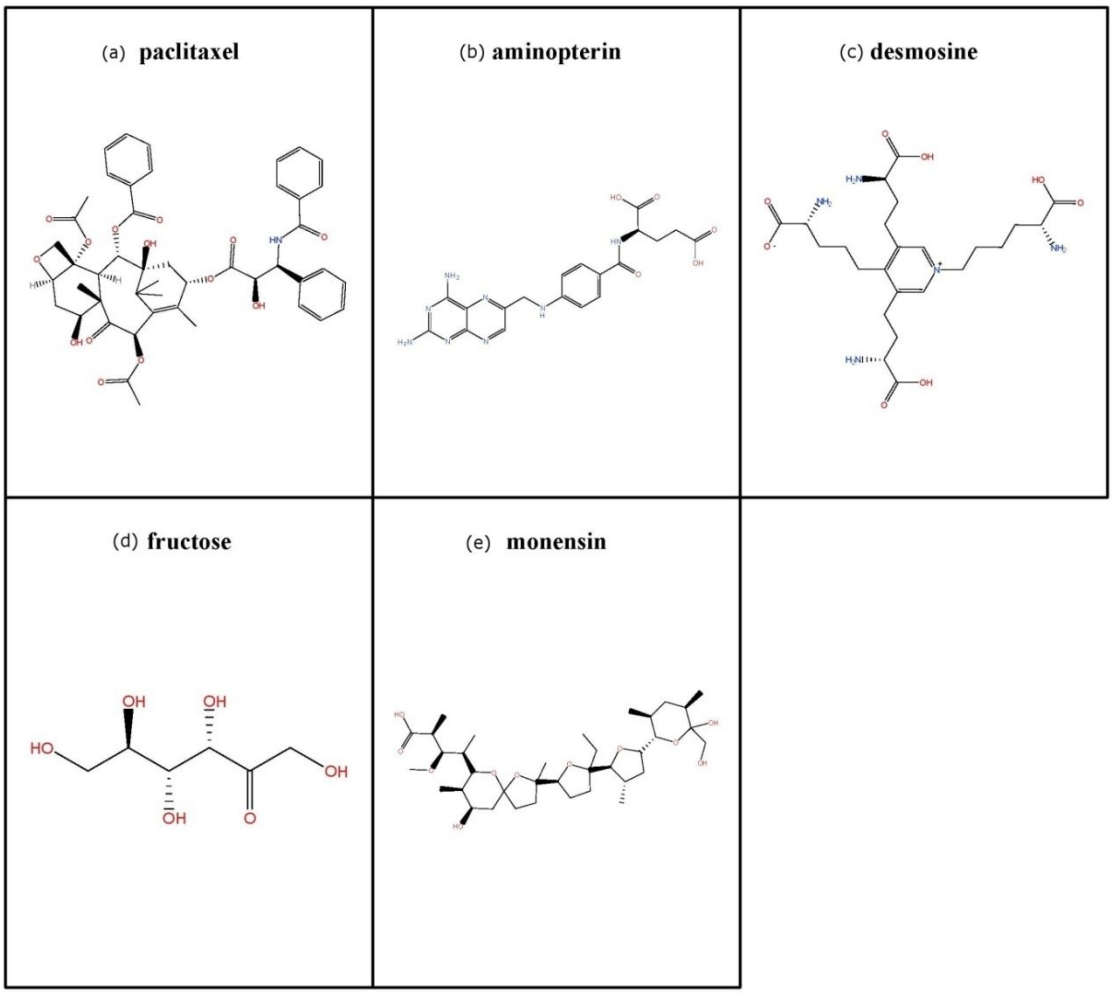


Figure 3.4 Chemical structures of the lead compounds discovered by Protein Scanning with Virtual Ligand Screening (PSVLS) combined with Holistic Binding scoring. These compounds were tested experimentally for inhibition and/or binding to BoNTA. (a) paclitaxel; (b) aminopterin; (c) desmosine; (d) fructose; (e) monensin.

Desmosine binds to the catalytic site of BoNTA with a score (-5.5 kcal/mol) just above our non-binder/binder threshold. This compound should display some inhibitory activity experimentally.

Aminopterin binds to the catalytic site of BoNTA with a score (-4.2 kcal/mol) above (worse than) the non-binder/binder threshold, but significantly lower (more favorable) than the

median value for the library (-1.9 kcal/mol). Hence, it is expected to exhibit low inhibitory activity, if any.

Monensin, a sodium and proton selective ionophore antibiotic, is known to delay the onset of botulism. Two mechanisms have been suggested for Monensin's ability to delay muscular paralysis: control of endosomal acidification which does not involve direct binding to the toxin, [69] and inhibition of toxin-induced channel formation.[70] In our calculations, monensin was found to bind to the catalytic site of BoNTA with a score of -4.1 kcal/mol. This value is above (worse than) the binder/non-binder threshold of -5 kcal/mol, but it is much better than the median (non-binder) value of -1.9 kcal/mol, similarly to aminopterin. We expect this compound to have low to no inhibitory activity.

Fructose has a score of -1.8 kcal/mol at the catalytic site, which places this compound in the "non-binder" category. Hence, this compound is expected to be inactive experimentally as inhibitor of BoNTA catalytic activity. The best score for fructose across all binding regions is -1.9 kcal/mol (region 17). Therefore, fructose is a good choice for negative control in the binding experiments targeting both whole BoNTA as well as BoNTA LC. The experimental confirmation of a non-binder is a good test of the prediction power of the overall methodology.

Table 3.1 Holistic Binding (HB) scores and predicted activity for inhibitor candidates selected for experimental testing. Scores were obtained from virtual screening of the *botulinum* neurotoxin subtype A (BoNTA), heavy and light chains, against 1,624 chemical compounds. A discussion of the rationale used for hit selection is presented in the text.

Ligand	HB Score (kcal/mol)	HB Ranking (catalytic site)[1]	HB Predicted Activity Binding to BoNTA LC with Inhibition of Catalytic Activity	FF Score (kcal/mol)	FF Ranking (catalytic site)[2]
Paclitaxel	-11.0	6	Yes (strong)	-90.9	91
D-(-)-fructose	-1.8	735	No	-29.1	1285
Desmosine	-5.5	28	May be (low to none)	-115.6	38
Aminopterin	-4.2	63	May be (low to none)	-118.6	36
Monensin	-4.1	70	May be (low to none)	-79.8	163

[1]Ranking based on Holistic Binding scores for the catalytic site of BoNTA. More negative scores are better. The threshold of binding energy used to consider a compound to be a "binder" in the catalytic site is -5 kcal/mol. Ligands with scores more negative than this threshold have better chance to bind experimentally to the target. Ranking reflects the inclusion of fluorescent derivatives of paclitaxel to the ranking list after paclitaxel was selected for testing and confirmed experimentally to bind to BoNTA.

[2]Ranking based on raw Force Field (FF) scores for the catalytic site of BoNTA, provided for comparison. The threshold of binding energy to consider a compound to be a "binder" in the catalytic site for FF-based scores is -106 kcal/mol. Among the compounds in this table, only aminopterin would be predicted as a "binder" by the FF-based scores. Using this scoring scheme, 5.6% of the ranked library would have to be tested experimentally to identify paclitaxel as a BoNTA inhibitor.

3.5.4 Inhibition of BoNTA catalytic activity by selected hits

The results of the FRET assay against the light chain (catalytic domain) of BoNTA are shown in Table 3.2. These preliminary experimental assays indicated that paclitaxel inhibits the botulism toxin's proteolytic activity experimentally by 95% at 10μM (semi-synthetic paclitaxel). Additional assays have found that paclitaxel isolated from natural sources inhibit BoNTA LC by 85% at 10μM. A dose-response curve, shown in Figure 3.5(a), was constructed for paclitaxel using the FRET assay. A four-parametric nonlinear regression model (Prism5, GraphPad Inc.)

determined the IC_{50} for paclitaxel to be 5.2 μM ± 1.1 μM over two independent experiments. The dose-response curve for EDTA (known to inhibit catalytic activity) is also shown in Figure 3.5(a) for comparison. Paclitaxel is solubilized using DMSO as co-solvent, which was found to be well tolerated in the SNAPtide assay for up to 25% DMSO.[71] Our results indicate that the fluorescence intensity produced by fixed concentrations of toxin and SNAPtide is stable up to 4.8% DMSO. At higher DMSO concentrations, the effect of the solvent needs to be taken into account. The concentration-dependency of paclitaxel binding to BoNTA LC was confirmed by a dose-response curve, shown in Figure 3.5(b), constructed using the SPA assay described earlier.

Table 3.2 Inhibition of BoNTA light chain catalytic activity by selected hits.

Ligand	% Inhibition at 10 μM[1]
Paclitaxel	95
D-(-)-fructose	0
Desmosine	25
Aminopterin	11
Monensin	38

[1]Percent inhibition was calculated in FRET assays using 8 μM of SNAPtide. The concentration of BoNTA LC recombinant protein was 10 nM. Assays were performed at 37˚C, pH 8.2, in triplicate. Toxin was added to solutions of buffer, substrate and ligands. Samples were incubated for 1 hour prior to readings.

(a) (b)

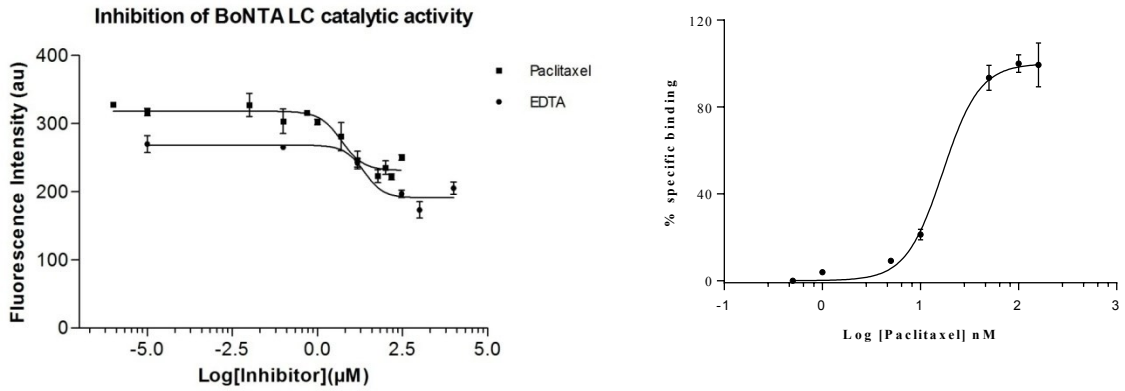


Figure 3.5 (a) Inhibition curve for paclitaxel and EDTA (positive control) against the FRET substrate SNAPtide® (8 μM). The IC50 for paclitaxel against the light chain (catalytic domain) of BoNTA (10 nM) was estimated to be 5.2 μM using a four-parametric nonlinear regression model (Prism5, GraphPad Inc.). (b) Specific binding curve for Paclitaxel [2-benzoyl ring-3H] against the light chain (catalytic domain) of BoNTA (7 nM) obtained in a SPA assay. Curve-fitting was performed with a four-parametric nonlinear regression model (Prism5, GraphPad Inc.). These results confirm the concentration-dependency of paclitaxel binding to BoNTA LC.

Paclitaxel is a marketed drug used for cancer treatment. It targets tubulin, interferes with the dynamic instability of microtubules and, thus, prevents cell division.[72] The use of paclitaxel as a therapeutic agent for botulism is likely to be undesirable because of its insolubility and its toxic effects. However, paclitaxel may provide a good starting point for the rational design of novel inhibitors for therapeutic use. Other terpene (farnesyl pyrophosphate -5.7 kcal/mol), terpene precursor (geranylgeranyl pyrophosphate -5.2 kcal/mol), and taxane (docetaxel -5.8 kcal/mol) were found to have favorable Holistic Binding scores in our calculations. Although they have not yet been confirmed experimentally, these compounds may provide additional assistance for rational design of potent BoNTA inhibitor. The finding that paclitaxel inhibits BoNTA may also have implications for the development of new anti-cancer treatments. Zinc deficiency was found to reduce paclitaxel efficacy in LNCaP prostate cancer cells.[73] This fact combined with our results for BoNTA inhibition suggests that paclitaxel inhibit zinc-dependent

proteases present in early-stage cancerous cells, in addition to its well-documented role as a disruptor of microtubule function. LNCaP cells are a model for early-stage prostate cancer and this additional role for paclitaxel may help in the identification of early-stage biomarkers for prostate cancer. The zinc-dependent effect was not observed in PC3 cell lines, which are a model for late-stage prostate cancer.[73]

As predicted by the analysis of Holistic Binding scores, fructose shows no inhibitory activity against BoNTA. The ligands predicted to be borderline for binding showed low inhibitory activity. Monensin displayed 38% inhibition at 10 μM, which is better than anticipated and suggests that direct inhibition may be a secondary mechanism to the ones already proposed for this ligand's ability to delay onset of botulism. [69-70] Desmosine and aminopterin showed low inhibitory activity (25% and 11% at 10μM), consistent with their Holistic Scores. Both are at the non-binder/binder threshold. Aminopterin showed intrinsic fluorescence at 100μM with emission peak at 463 nm, and may interfere with fluorescence readings at high concentrations.

The force field binding energies (FF_BE) and ranking for each one of the compounds discussed above is also provided in Table 3.1 for comparison. The non-binder/binder threshold for the catalytic region of BoNTA using FF_BE values is -106 kcal/mol. Under the FF_BE ranking, only aminopterin would have been considered a "binder", and the identification of paclitaxel would have required the experimental testing of 5.6% of the screening library. These results emphasize the better performance of the holistic binding scheme compared to traditional force field-based scoring.

3.5.5 Paclitaxel fluorescent derivatives

In addition to the search for novel potent inhibitors, fast and reliable detection of BoNTA is also of concern. The standard test for detection of BoNT is still the mouse lethality assay.

[74]This assay is laborious, expensive, time-consuming, and ethically controversial. Other technologies are available [74-80] but none is ideal for fast, low cost and portable devices that accurately detect the presence of the toxin in a variety of field, clinical and food samples. Existing detection technologies are based on the catalytic activity of BoNTA. However, the BoNTA catalytic site is believed to be occluded when the light and heavy chain are connected. [67]Since the toxin only releases its light chain under reducing conditions, the efficacy of these assays depends on the breakage of the disulfide bond connecting the light and heavy chains. Therefore, the discovery of small molecules suitable as recognition agents for BoNTA is desirable.

Upon experimental confirmation of paclitaxel as an inhibitor of BoNTA, we investigated fluorescent derivatives of paclitaxel that could be used as recognition agents for detection of BoNTA. Three commercially available derivatives were tested experimentally for binding to whole BoNTA. They were tested against whole BoNTA because the catalytic site of BoNTA is believed to be partially occluded in whole BoNTA, and a good recognition agent must be able to detect the whole toxin. This is the form of the toxin before the catalytic domain is translocated with the help of the translocation domain into the cytosol, where neurotransmission is inhibited as a result of SNARE protein cleavage by the catalytic domain.[67] The paclitaxel fluorescent derivatives were also submitted to PSVLS/Holistic Binding scoring. The intention of these calculations was to verify whether the Holistic Binding approach could have identified the experimentally positive compounds as "good binders". All three derivatives scored well (Table 3.3), but only one of them was found to effectively bind to whole BoNTA (Table 3.3). Analysis of the ligand-protein interactions for the best PSVLS complex of each paclitaxel fluorescent derivative revels that hydrogen bonds to Gln162 are a common feature. The paclitaxel moiety is

buried in the binding site in the structures of pac-BDP and pac-FL but more exposed to solvent in the pac-OG structure. The fluorophore (BDP, FL or OG moiety) is exposed in the structure of pac-BDP, somewhat exposed in pac-FL and buried within the binding cavity in pac-OG. This fluorescence intensity assay results against whole BoNTA follow the same trend. Hydrogen bonds to Gln162, Lys166, Glu64 and Arg536 anchors pac-OG to the site and are likely responsible for a more favorable calculated binding affinity compared to pac-BDP and pac-FL. The ligand-protein interaction diagrams are available as supporting information.

These fluorescent paclitaxel derivatives are used for imaging microtubule formation and for labeling microtubules in solution. Their direct use as BoNTA recognition agent is limited, as their known target (tubulin) could cause a false positive reading in the assays. However, the experimentally confirmed conjugate (pac-BDP) can be used as a lead for the design of highly-specific fluorescent probes for BoNTA. This compound can also be useful in multi-probe fluorescence-based detection assays, where elimination of false positives can be accomplished by means of combinatorial response.

Table3.3 Binding of fluorescent recognition agent candidates against the *botulinum* neurotoxin subtype A non-activated heavy and light chain complex (whole BoNTA). Binding to the catalytic domain (light chain) of BoNTA and inhibition of catalytic activity does not assure binding to whole BoNTA, which is necessary for reliable detection.

Ligand	Holistic Binding Score at the Catalytic Site (Ranking[1]) (kcal/mol)	Experimental Binding Against Non-activated Whole BoNTA (%Bound Ligand)	
		Fluorescence Intensity Assay[2]	Radiolabeled Ligand Binding Assay[3]
D-(-)-fructose	-1.8 (735)		3
Paclitaxel (Pa)	-11.0 (6)		7
Pac-OG	-22.0 (1)	4	
Pac-BDP	-16.0 (2)	46	
Pac-FL	-12.4 (3)	22	

[1]Ranking based on Holistic Binding scores for the catalytic and for the best non-catalytic binding site of BoNTA. More negative scores are better. Derivatives were added to the ranking list after paclitaxel was selected for testing and confirmed experimentally to bind to BoNTA. All ligands were re-ranked after inclusion of paclitaxel fluorescent derivatives to the ranking list.

[3]Radiolabeled Ligand Binding. Paclitaxel and fructose were tested in duplicate at 50 μM (paclitaxel[14C] and fructose[14C]), 25 μM (paclitaxel[14C]) and 10 μM (fructose[14C]) for binding to non-activated 0.1 μg/ml whole BoNTA, at room temperature and two pHs (5.5 and 7.4), with consistent results. Biotinylated toxin was attached to streptavidin-coated magnetic beads for separation of free and bound ligands.

[2]Fluorescence Intensity (FI). Fluorescent ligands were tested at 10 μM against non-activated 0.1 μg/ml whole BoNTA. The toxin was attached to magnetic beads for separation. FI was measured in duplicate relative to the reference sample of each fluorescent ligand, after 1 hour incubation at room temperature and pH 7.4. Excitation/Emission wavelengths: Pac-FL (505nm/515nm); Pac-BDP (565nm/571nm); Pac-OG (496nm/524nm).

3.5.6 Binding of paclitaxel to whole BoNTA

Inhibitors unable to bind to BoNTA before the catalytic domain reaches the cytosol may be less effective than inhibitors that target whole BoNTA. In order to probe whether paclitaxel binds to whole BoNTA in addition to the exposed BoNTA LC, radiolabeled ligand binding experiments were carried out for ^{14}C-paclitaxel and ^{14}C-fructose against non-activated BoNTA

complex. Fructose was used as negative control as it is not expected to bind to any form of BoNTA. The toxin was attached to magnetic beads for separation. Absorption at 280 nm was used to confirm that the beads were loaded with toxin prior to the assay. Paclitaxel was tested at 50 μM and 25 μM, while fructose was tested at 50 μM and 10 μM. The amount of toxin was kept constant. Specific binding of paclitaxel to the whole toxin was consistent at both concentrations (7%). Specific binding of fructose to BoNTA was 5% at 10 μM and 3% at 50 μM. Binding to non-loaded beads (non-specific binding) was 29% (25 μM) and 23% (50 μM) for paclitaxel, and 3% for fructose at both concentrations. The low solubility of paclitaxel in water is likely responsible for this relatively high non-specific binding. These results (Table 3.3) indicate that paclitaxel does not bind to whole BoNTA, even though it inhibits the catalytic activity of the toxin as confirmed by the FRET assay. However, further tests are necessary to eliminate the possibility that the immobilization of the toxin prevented the binding of paclitaxel by making its binding site less accessible. In addition, full biochemical characterization of this previously unidentified inhibitor is necessary to assess its potential as therapeutic agent for botulism.

3.6 CONCLUSION

We used the Holistic Binding computer-assisted approach to successfully identify a previously unknown inhibitor of BoNTA catalytic activity, paclitaxel. We showed experimentally that paclitaxel binds (SPA assay) and inhibits (FRET assay) BoNTA light chain but does not bind to whole BoNTA (radiolabeled ligand binding assay), while its fluorescent derivative pac-BDP binds to the heavy and light chain BoNTA complex (fluorescence intensity assay).

The predictions from Holistic Binding are consistent with experimental observations, supporting the use of this technique for discovery of novel inhibitors and recognition agents for

BoNTA. The concept of non-binder/binder thresholds determined by the mean and standard deviation values for the virtual library was shown to be an effective way to select hits for experimental testing in the case of BoNTA. Further evaluation of false positive and false negative rates is needed to adequately determine its predictive power. Nonetheless, we expect that non-binder/binder thresholds will be useful in the context of other virtual screening approaches as well.

The success rate of a VLS run largely depends on the size and diversity of the chemical library,[81] and on the fortuitous presence of true binders (positives). In general, more is better and the screening of large libraries has a better chance to yield true binders. Our screening of a small library may seem to have been more successful than one could have expected. However, based on reported success rates for first-tier High Throughput Screening (0.2% to 1.91%), [65, 82-83] a library of the size we virtually screened might have 3-31 positive compounds. The ability to enrich the top of a ranked list with whatever true binders are present in the library is the main challenge of VLS. The *HierVLS* method used at the core of PSVLS was shown to yield good enrichment factors. [47] Our present results, albeit not conclusive, indicate that the Holistic Binding approach is efficient at concentrating true binders at the top of the ranked list.

The Holistic Binding approach is a time- and cost-efficient alternative to other discovery approaches, such as HTS. For example, a recent HTS[65] of 16,544 compounds found 34 BoNTA inhibitors in the first-tier screening (0.2% hit rate), which were reduced to 9 inhibitors upon a confirmatory ELISA assay (0.05% hit rate). The best inhibitor in this group showed 95% inhibition of the toxin in a FRET assay with 50nM BoNTA and 5μM SNAPTide® substrate. In contrast, our virtual screening of a library ten times smaller readily identified an inhibitor with comparable performance in a similar FRET assay. Although the fortuitous presence of active

ligands in the libraries being (experimentally or computationally) screened is certainly a determinant of success, our methodology was shown to be effective at identifying active ligands at a more accessible cost. VLS generally requires expensive High Performance Computer Clusters (HPCCs). However, the recent availability of service-based cloud computing environments such as Amazon Elastic Compute Cloud (EC2) (aws.amazon.com/ec2/) is expected to make virtual screening more affordable.

The need for novel fluorescent recognition agents for BoNTA is reinforced by the recent development of a multi-wavelength spatial LED illumination based detector that utilizes SNAPtide® as probe.[84] The detector technology for a fluorescent platform is already in place, but its applicability is curtailed by the limitations associated to the available probes. Although reliable, SNAPTide® only works if the catalytic domain of the toxin is activated and fully available for binding. The development of small-molecule probes capable of working in a wide range of conditions is, thus, desirable and will have immediate application. Within this goal, pac-BDP is currently being used by our group as a lead compound for the structure-based design of new chemical entities with improved affinity and selectivity for BoNTA. The existence of an experimentally determined structure for paclitaxel-tubulin (pdb codes 2HXF, 2HXH and 1IA0) along with our predicted ligand-BoNTA structure enables the rational design of analogs with reduced affinity for tubulin and increased affinity for BoNTA. The same strategy is being applied for the design of novel BoNTA inhibitors based on paclitaxel.

3.7 ASSOCIATED CONTENT

Supporting information. The following data are provided as supplementary material in a single pdf file: 2D ligand-protein interaction diagrams for (a) paclitaxel, (b) pac-BDP, (c) pac-FL, and (d) pac-OG bound to BoNTA, force field and Holistic Scores for each ligand in the

screening library docked to each of 33 binding sites in the target protein, and force field and Holistic Scores for the best conformer per ligand found to be in the catalytic site of the target protein. This material is available free of charge at http://pubs.acs.org.

PDB codes. 2NYY, 3BTA

3.8 AUTHOR INFORMATION

Corresponding author. Wely Floriano, Lakehead University, Department of Chemistry, 955 Oliver Rd Thunder Bay ON P7B 5E1, Canada, Email: wely.floriano@lakeheadu.ca, Phone: (807) 766-7215.

Present addresses.Zack Ramjan, USC Epigenome Center, 1450 Biggy Street Building Code: NRT G517C, Los Angeles CA 90033-9601, Email: ramjan@usc.edu, Phone: (323) 442-7932

3.9 ACKNOWLEDGEMENTS

This research was supported with funds from the National Institutes of Health (NIH) (S06 GM053933). We thank Dr. Sepehr Eskandari for invaluable advice regarding the radiolabeled ligand binding experiments, Andrew P. Clark for assisting with data manipulation, and Desiree Daniels for assisting with the preparation of solutions and microplates. Part of this work utilized facilities and equipment available at the Biological Sciences Department, California State Polytechnic University Pomona, the corresponding author's previous affiliation.

3.10 ABBREVIATIONS USED

BoNT, *Botulinum* neurotoxin; BoNTA, *Botulinum* Neurotoxin Subtype A; CAS, Chemical Abstracts Service; EF, Enrichment Factor; FF_BE, force field-based binding energies; HBS,

Holistic Binding Score; HC, heavy chain; *HierVLS,* Hierarchical Virtual Ligand Screening; LED, light-emitting diode; LC, light chain; NCIOR, National Cancer Institute Open Repository; Pa-OG, paclitaxel, 2',7'-difluorofluorescein conjugate; Pa-BDP, paclitaxel, 4,4-difluoro-5-styryl-4-bora-3a,4a-diaza-s-indacene-3-propionic acid conjugate; Pa-FL, paclitaxel, 4,4-difluoro-5,7-dimethyl-4-bora-3a,4a-diaza-s-indacene-3-propionic acid conjugate; PE, potential energy; PSVLS, Protein Scanning with Virtual Ligand Screening; RLB, Radiolabeled Ligand Binding; SE, solvation energy; SNARE, specific soluble N-ethylmaleimide-sensitive attachment protein receptor; SNAP-25, synaptosome-associated protein of 25 kDa; VLS, Virtual Ligand Screening.

3.11 REFERENCES

1. Arnon, S. S.; Schechter, R.; Inglesby, T. V.; Henderson, D. A.; Bartlett, J. G.; Ascher, M. S.; Eitzen, E.; Fine, A. D.; Hauer, J.; Layton, M.; Lillibridge, S.; Osterholm, M. T.; O'Toole, T.; Parker, G.; Perl, T. M.; Russell, P. K.; Swerdlow, D. L.; Tonat, K. Botulinum toxin as a biological weapon: medical and public health management. *JAMA* **2001,** 285, 1059-1070.

2. Greenfield, R. A.; Brown, B. R.; Hutchins, J. B.; Iandolo, J. J.; Jackson, R.; Slater, L. N.; Bronze, M. S. Microbiological, biological, and chemical weapons of warfare and terrorism. *Am J Med Sci* **2002,** 323, 326-340.

3. Sobel, J.; Tucker, N.; Sulka, A.; McLaughlin, J.; Maslanka, S. Foodborne botulism in the United States, 1990-2000. *Emerg Infect Dis* **2004,** 10, 1606-1611.

4. Yamasaki, S.; Baumeister, A.; Binz, T.; Blasi, J.; Link, E.; Cornille, F.; Roques, B.; Fykse, E. M.; Sudhof, T. C.; Jahn, R.; Niemann, H. Cleavage of members of the synaptobrevin/VAMP family by types D and F botulinal neurotoxins and tetanus toxin. *J Biol Chem* **1994,** 269, 12764-12772.

5. Yamasaki, S.; Binz, T.; Hayashi, T.; Szabo, E.; Yamasaki, N.; Eklund, M.; Jahn, R.; Niemann, H. Botulinum neurotoxin type G proteolyses the Ala81-Ala82 bond of rat synaptobrevin 2. *Biochem Biophys Res Commun* **1994,** 200, 829-835.

6. Binz, T.; Blasi, J.; Yamasaki, S.; Baumeister, A.; Link, E.; Sudhof, T. C.; Jahn, R.; Niemann, H. Proteolysis of SNAP-25 by types E and A botulinal neurotoxins. *J Biol Chem* **1994,** 269, 1617-1620.

7. Blasi, J.; Binz, T.; Yamasaki, S.; Link, E.; Niemann, H.; Jahn, R. Inhibition of neurotransmitter release by clostridial neurotoxins correlates with specific proteolysis of synaptosomal proteins. *J Physiol Paris* **1994,** 88, 235-241.

8. Schiavo, G.; Benfenati, F.; Poulain, B.; Rossetto, O.; Polverino de Laureto, P.; DasGupta, B. R.; Montecucco, C. Tetanus and botulinum-B neurotoxins block neurotransmitter release by proteolytic cleavage of synaptobrevin. *Nature* **1992,** 359, 832-835.

9. Washbourne, P.; Pellizzari, R.; Baldini, G.; Wilson, M. C.; Montecucco, C. Botulinum neurotoxin types A and E require the SNARE motif in SNAP-25 for proteolysis. *Febs Letters* **1997,** 418, 1-5.

10. Shukla, H. D.; Sharma, S. K. Clostridium botulinum: a bug with beauty and weapon. *Crit Rev Microbiol* **2005,** 31, 11-18.

11. Dolly, J. O.; Lawrence, G. W.; Meng, J. H.; Wang, J. F.; Ovsepian, S. V. Neuro-exocytosis: botulinum toxins as inhibitory probes and versatile therapeutics. *Current Opinion in Pharmacology* **2009,** 9, 326-335.

12. Koussoulakos, S. Botulinum Neurotoxin: The Ugly Duckling. *European Neurology* **2009,** 61, 331-342.

13. Stone, C. A.; O'Leary, N. Systematic Review of the Effectiveness of Botulinum Toxin or Radiotherapy for Sialorrhea in Patients with Amyotrophic Lateral Sclerosis. *Journal of Pain and Symptom Management* **2009,** 37, 246-258.

14. Restivo, D. A.; Marchese-Ragona, R.; Lauria, G.; Squatrito, S.; Gullo, D.; Vigneri, R. Botulinum toxin treatment for oropharyngeal dysphagia associated with diabetic neuropathy. *Diabetes Care* **2006,** 29, 2650-2653.

15. Schulte-Mattler, W. J. Use of botulinum toxin A in adult neurological disorders - Efficacy, tolerability and safety. *Cns Drugs* **2008,** 22, 725-738.

16. Jankovic, J. Botulinum toxin in clinical practice. *Journal of Neurology Neurosurgery and Psychiatry* **2004,** 75, 951-957.

17. Brisinda, G.; Cadeddu, F.; Vanella, S.; Mazzeo, P.; Maria, G. Relief by Botulinum Toxin of Lower Urinary Tract Symptoms Owing to Benign Prostatic Hyperplasia: Early and Long-Term Results. *Urology* **2009,** 73, 90-94.

18. Restivo, D. A.; Palmeri, A.; Marchese-Ragona, R. Botulinum toxin for cricopharyngeal dysfunction in Parkinson's disease. *New England Journal of Medicine* **2002,** 346, 1174-1175.

19. Aoki, K. R. Pharmacology and immunology of botulinum toxin serotypes. *J Neurol* **2001,** 248 Suppl 1, 3-10.

20. Colosimo, C.; Salvatori, F. M. Injection of the iliopsoas muscle with botulinum toxin in camptocormia. *Mov Disord* **2009,** 24, 316-317.

21. von Coelln, R.; Raible, A.; Gasser, T.; Asmus, F. Ultrasound-guided injection of the iliopsoas muscle with botulinum toxin in camptocormia. *Mov Disord* **2008,** 23, 889-892.

22. Hexsel, D.; Spencer, J. M.; Woolery-Lloyd, H.; Gilbert, E. Practical Applications of a New Botulinum Toxin. *Journal of Drugs in Dermatology* **2010,** 9, S31-S37.

23. ASAPS. 15th Annual Cosmetic Surgery National Data Bank Statistics. In March 20, 2012 ed.; The American Society for Aesthetic Plastic Surgery (ASAPS): 2012.

24. De Bouelle, K. L. V. Botulinum neurotoxin type A in facial aesthetics. *Expert Opinion on Pharmacotherapy* **2007,** 8, 1059-1072.

25. ISAPS. ISAPS International Survey on Aesthetic/Cosmetic Procedures Performed in 2011. In International Society of Aesthetic Plastic Surgery (ISAPS): 2012.

26. Arnold, P. B.; Merritt, W.; Rodeheaver, G. T.; Campbell, C. A.; Morgan, R. F.; Drake, D. B. Effects of perivascular Botulinum Toxin-A application on vascular smooth muscle and flap viability in the rat. *Ann Plast Surg* **2009,** 62, 463-467.

27. Roche, N.; Schnitzler, A.; Genet, F. F.; Durand, M. C.; Bensmail, D. Undesirable distant effects following botulinum toxin type a injection. *Clin Neuropharmacol* **2008,** 31, 272-280.

28. Curra, A.; Berardelli, A. Do the unintended actions of botulinum toxin at distant sites have clinical implications? *Neurology* **2009,** 72, 1095-1099.

29. Antonucci, F.; Rossi, C.; Gianfranceschi, L.; Rossetto, O.; Caleo, M. Long-distance retrograde effects of botulinum neurotoxin A. *Journal of Neuroscience* **2008,** 28, 3689-3696.

30. Lange, D. J.; Rubin, M.; Greene, P. E.; Kang, U. J.; Moskowitz, C. B.; Brin, M. F.; Lovelace, R. E.; Fahn, S. Distant effects of locally injected botulinum toxin: a double-blind study of single fiber EMG changes. *Muscle Nerve* **1991,** 14, 672-675.

31. Sheffield, J. K.; Jankovic, J. Botulinum toxin in the treatment of tremors, dystonias, sialorrhea and other symptoms associated with Parkinson's disease. *Expert Rev Neurother* **2007,** 7, 637-47.

32. Rosales, R. L.; Bigalke, H.; Dressler, D. Pharmacology of botulinum toxin: differences between type A preparations. *Eur J Neurol* **2006,** 13 Suppl 1, 2-10.

33. Schroeder, A. S.; Ertl-Wagner, B.; Britsch, S.; Schroder, J. M.; Nikolin, S.; Weis, J.; Muller-Felber, W.; Koerte, I.; Stehr, M.; Berweck, S.; Borggraefe, I.; Heinen, F. Muscle biopsy substantiates long-term MRI alterations one year after a single dose of botulinum toxin injected into the lateral gastrocnemius muscle of healthy volunteers. *Mov Disord* **2009,** 24, 1494-1503.

34. Krystkowiak, P.; Kreisler, A. [High doses of botulinum toxin, from efficacy to safety: experimental data]. *Ann Readapt Med Phys* **2007,** 50 Suppl 1, S12-S16.

35. Eubanks, L. M.; Dickerson, T. J.; Janda, K. D. Technological advancements for the detection of and protection against biological and chemical warfare agents. *Chem Soc Rev* **2007,** 36, 458-470.

36. Burnett, J. C.; Opsenica, D.; Sriraghavan, K.; Panchal, R. G.; Ruthel, G.; Hermone, A. R.; Nguyen, T. L.; Kenny, T. A.; Lane, D. J.; McGrath, C. F.; Schmidt, J. J.; Vennerstrom, J. L.; Gussio, R.; Solaja, B. A.; Bavari, S. A refined pharmacophore identifies potent 4-amino-7-chloroquinoline-based inhibitors of the botulinum neurotoxin serotype A metalloprotease. *J Med Chem* **2007,** 50, 2127-2136.

37. Burnett, J. C.; Ruthel, G.; Stegmann, C. M.; Panchal, R. G.; Nguyen, T. L.; Hermone, A. R.; Stafford, R. G.; Lane, D. J.; Kenny, T. A.; McGrath, C. F.; Wipf, P.; Stahl, A. M.; Schmidt, J. J.; Gussio, R.; Brunger, A. T.; Bavari, S. Inhibition of metalloprotease botulinum serotype A from a pseudo-peptide binding mode to a small molecule that is active in primary neurons. *J Biol Chem* **2007,** 282, 5004-5014.

38. Burnett, J. C.; Wang, C.; Nuss, J. E.; Nguyen, T. L.; Hermone, A. R.; Schmidt, J. J.; Gussio, R.; Wipf, P.; Bavari, S. Pharmacophore-guided lead optimization: the rational design of a non-zinc coordinating, sub-micromolar inhibitor of the botulinum neurotoxin serotype a metalloprotease. *Bioorg Med Chem Lett* **2009,** 19, 5811-5813.

39. Capkova, K.; Salzameda, N. T.; Janda, K. D. Investigations into small molecule non-peptidic inhibitors of the botulinum neurotoxins. *Toxicon* **2009,** 54, 575-582.

40. Eubanks, L. M.; Hixon, M. S.; Jin, W.; Hong, S.; Clancy, C. M.; Tepp, W. H.; Baldwin, M. R.; Malizio, C. J.; Goodnough, M. C.; Barbieri, J. T.; Johnson, E. A.; Boger, D. L.; Dickerson, T. J.; Janda, K. D. An in vitro and in vivo disconnect uncovered through high-throughput identification of botulinum neurotoxin A antagonists. *Proc Natl Acad Sci U S A* **2007,** 104, 2602-2607.

41. Hermone, A. R.; Burnett, J. C.; Nuss, J. E.; Tressler, L. E.; Nguyen, T. L.; Solaja, B. A.; Vennerstrom, J. L.; Schmidt, J. J.; Wipf, P.; Bavari, S.; Gussio, R. Three-dimensional database

mining identifies a unique chemotype that unites structurally diverse botulinum neurotoxin serotype A inhibitors in a three-zone pharmacophore. *ChemMedChem* **2008,** 3, 1905-1912.

42. Park, J. G.; Sill, P. C.; Makiyi, E. F.; Garcia-Sosa, A. T.; Millard, C. B.; Schmidt, J. J.; Pang, Y. P. Serotype-selective, small-molecule inhibitors of the zinc endopeptidase of botulinum neurotoxin serotype A. *Bioorg Med Chem* **2006,** 14, 395-408.

43. Zuniga, J. E.; Schmidt, J. J.; Fenn, T.; Burnett, J. C.; Arac, D.; Gussio, R.; Stafford, R. G.; Badie, S. S.; Bavari, S.; Brunger, A. T. A potent peptidomimetic inhibitor of botulinum neurotoxin serotype A has a very different conformation than SNAP-25 substrate. *Structure* **2008,** 16, 1588-1597.

44. Capkova, K.; Hixon, M. S.; Pellett, S.; Barbieri, J. T.; Johnson, E. A.; Janda, K. D. Benzylidene cyclopentenediones: First irreversible inhibitors against botulinum neurotoxin A's zinc endopeptidase. *Bioorg Med Chem Lett* **2010,** 20, 206-208.

45. Dembek, Z. F.; Smith, L. A.; Rusnak, J. M. Botulism: cause, effects, diagnosis, clinical and laboratory identification, and treatment modalities. *Disaster Med Public Health Prep* **2007,** 1, 122-134.

46. Pang, Y. P.; Davis, J.; Wang, S.; Park, J. G.; Nambiar, M. P.; Schmidt, J. J.; Millard, C. B. Small molecules showing significant protection of mice against botulinum neurotoxin serotype A. *PLoS One* **2010,** 5, e10129.

47. Floriano, W. B.; Vaidehi, N.; Zamanakos, G.; Goddard, W. A., 3rd. HierVLS hierarchical docking protocol for virtual ligand screening of large-molecule databases. *J Med Chem* **2004,** 47, 56-71.

48. Ramjan, Z. H.; Raheja, A.; Floriano, W. B. A cluster-aware graphical user interface for a virtual ligand screening tool. *Conf Proc IEEE Eng Med Biol Soc* **2008,** 2008, 4102-4105.

49. Brunger, A. T.; Breidenbach, M. A.; Jin, R.; Fischer, A.; Santos, J. S.; Montal, M. Botulinum neurotoxin heavy chain belt as an intramolecular chaperone for the light chain. *PLoS Pathog* **2007,** 3, 1191-1194.

50. Breidenbach, M. A.; Brunger, A. T. Substrate recognition strategy for botulinum neurotoxin serotype A. *Nature* **2004,** 432, 925-929.

51. Garcia-Rodriguez, C.; Levy, R.; Arndt, J. W.; Forsyth, C. M.; Razai, A.; Lou, J.; Geren, I.; Stevens, R. C.; Marks, J. D. Molecular evolution of antibody cross-reactivity for two subtypes of type A botulinum neurotoxin. *Nat Biotechnol* **2007,** 25, 107-116.

52. Mayo, S. L.; Olafson, B. D.; Goddard, W. A. Dreiding - a Generic Force-Field for Molecular Simulations. *Journal of Physical Chemistry* **1990,** 94, 8897-8909.

53. MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *Journal of Physical Chemistry B* **1998,** 102, 3586-3616.

54. Brady, G. P., Jr.; Stouten, P. F. Fast prediction and visualization of protein binding pockets with PASS. *J Comput Aided Mol Des* **2000,** 14, 383-401.

55. Brylinski, M.; Skolnick, J. A threading-based method (FINDSITE) for ligand-binding site prediction and functional annotation. *Proc Natl Acad Sci U S A* **2008,** 105, 129-134.

56. Coleman, R. G.; Sharp, K. A. Protein pockets: inventory, shape, and comparison. *J Chem Inf Model* **2010,** 50, 589-603.

57. Laurie, A. T.; Jackson, R. M. Q-SiteFinder: an energy-based method for the prediction of protein-ligand binding sites. *Bioinformatics* **2005,** 21, 1908-1916.

58. Morris, R. J.; Najmanovich, R. J.; Kahraman, A.; Thornton, J. M. Real spherical harmonic expansion coefficients as 3D shape descriptors for protein binding pocket and ligand comparisons. *Bioinformatics* **2005,** 21, 2347-2355.

59. Oda, A.; Yamaotsu, N.; Hirono, S. Evaluation of the searching abilities of HBOP and HBSITE for binding pocket detection. *J Comput Chem* **2009,** 30, 2728-2737.

60. Ren, J.; Xie, L.; Li, W. W.; Bourne, P. E. SMAP-WS: a parallel web service for structural proteome-wide ligand-binding site comparison. *Nucleic Acids Res* **2010,** 38, W441-W444.

61. Tseng, Y. Y.; Chen, Z. J.; Li, W. H. fPOP: footprinting functional pockets of proteins by comparative spatial patterns. *Nucleic Acids Res* **2010,** 38, D288-D295.

62. Weisel, M.; Proschak, E.; Schneider, G. PocketPicker: analysis of ligand binding-sites with shape descriptors. *Chem Cent J* **2007,** 1, 7.

63. Xiong, B.; Wu, J.; Burk, D. L.; Xue, M.; Jiang, H.; Shen, J. BSSF: a fingerprint based ultrafast binding site similarity search and function analysis server. *BMC Bioinformatics* **2010,** 11, 47.

64. Zhang, Z.; Li, Y.; Lin, B.; Schroeder, M.; Huang, B. Identification of cavities on protein surface using multiple computational approaches for drug binding site prediction. *Bioinformatics* **2011,** 27, 2083-2088.

65. Cai, S.; Lindo, P.; Park, J. B.; Vasa, K.; Singh, B. R. The identification and biochemical characterization of drug-like compounds that inhibit botulinum neurotoxin serotype A endopeptidase activity. *Toxicon* **2010,** 55, 818-826.

66. Schiavo, G.; Rossetto, O.; Catsicas, S.; Polverino de Laureto, P.; DasGupta, B. R.; Benfenati, F.; Montecucco, C. Identification of the nerve terminal targets of botulinum neurotoxin serotypes A, D, and E. *J Biol Chem* **1993,** 268, 23784-23787.

67. Lacy, D. B.; Stevens, R. C. Sequence homology and structural analysis of the clostridial neurotoxins. *J Mol Biol* **1999,** 291, 1091-1104.

68. Fujinaga, Y. Transport of bacterial toxins into target cells: pathways followed by cholera toxin and botulinum progenitor toxin. *J Biochem* **2006,** 140, 155-160.

69. Sheridan, R. E. Protonophore antagonism of botulinum toxin in mouse muscle. *Toxicon* **1996,** 34, 849-855.

70. Deshpande, S. S.; Sheridan, R. E.; Adler, M. Efficacy of certain quinolines as pharmacological antagonists in botulinum neurotoxin poisoning. *Toxicon* **1997,** 35, 433-445.

71. Boldt, G. E.; Kennedy, J. P.; Hixon, M. S.; McAllister, L. A.; Barbieri, J. T.; Tzipori, S.; Janda, K. D. Synthesis, characterization and development of a high-throughput methodology for the discovery of botulinum neurotoxin a inhibitors. *J Comb Chem* **2006,** 8, 513-521.

72. Mitra, A.; Sept, D. Taxol allosterically alters the dynamics of the tubulin dimer and increases the flexibility of microtubules. *Biophys J* **2008,** 95, 3252-3258.

73. Killilea, A. N.; Downing, K. H.; Killilea, D. W. Zinc deficiency reduces paclitaxel efficacy in LNCaP prostate cancer cells. *Cancer Letters* **2007,** 258, 70-79.

74. Lindstrom, M.; Korkeala, H. Laboratory diagnostics of botulism. *Clin Microbiol Rev* **2006,** 19, 298-314.

75. Marazuela, D.; Moreno-Bondi, M. C. Fiber-optic biosensors--an overview. *Anal Bioanal Chem* **2002,** 372, 664-682.

76. Iqbal, S. S.; Mayo, M. W.; Bruno, J. G.; Bronk, B. V.; Batt, C. A.; Chambers, J. P. A review of molecular recognition technologies for detection of biological threat agents. *Biosens Bioelectron* **2000,** 15, 549-578.

77. Hallis, B.; James, B. A.; Shone, C. C. Development of novel assays for botulinum type A and B neurotoxins based on their endopeptidase activities. *J Clin Microbiol* **1996,** 34, 1934-1938.

78. Boyer, A. E.; Moura, H.; Woolfitt, A. R.; Kalb, S. R.; McWilliams, L. G.; Pavlopoulos, A.; Schmidt, J. G.; Ashley, D. L.; Barr, J. R. From the mouse to the mass spectrometer: detection and differentiation of the endoproteinase activities of botulinum neurotoxins A-G by mass spectrometry. *Anal Chem* **2005,** 77, 3916-3924.

79. Hines, H. B.; Kim, A. D.; Stafford, R. G.; Badie, S. S.; Brueggeman, E. E.; Newman, D. J.; Schmidt, J. J. Use of a recombinant fluorescent substrate with cleavage sites for all botulinum neurotoxins in high-throughput screening of natural product extracts for inhibitors of serotypes A, B, and E. *Appl Environ Microbiol* **2008,** 74, 653-659.

80. Schmidt, J. J.; Stafford, R. G.; Millard, C. B. High-throughput assays for botulinum neurotoxin proteolytic activity: serotypes A, B, D, and F. *Anal Biochem* **2001,** 296, 130-137.

81. Scior, T.; Bender, A.; Tresadern, G.; Medina-Franco, J. L.; Martinez-Mayorga, K.; Langer, T.; Cuanalo-Contreras, K.; Agrafiotis, D. K. Recognizing Pitfalls in Virtual Screening: A Critical Review. *Journal of Chemical Information and Modeling* **2012,** 52, 867-881.

82. Gribbon, P.; Lyons, R.; Laflin, P.; Bradley, J.; Chambers, C.; Williams, B. S.; Keighley, W.; Sewing, A. Evaluating real-life high-throughput screening data. *Journal of Biomolecular Screening* **2005,** 10, 99-107.

83. Shun, T. Y.; Lazo, J. S.; Sharlow, E. R.; Johnston, P. A. Identifying Actives from HTS Data Sets: Practical Approaches for the Selection of an Appropriate HTS Data-Processing Method and Quality Control Review. *Journal of Biomolecular Screening* **2011,** 16, 1-14.

84. Sun, S.; Ossandon, M.; Kostov, Y.; Rasooly, A. Lab-on-a-chip for botulinum neurotoxin a (BoNT-A) activity analysis. *Lab on a Chip* **2009,** 9, 3275-3281.

4 CHAPTER FOUR: TOWARDS A MULTI-PROBES FLUORESCENCE-BASED ASSAY AGAINST BoNTA WITH NOVEL FLUORESCENT MOLECULAR PROBES

*Saedeh Dadgar, Morshed Chowdhury, Christopher P. Phenix and Wely B. Floriano**

To be submitted to the *J. Med. Chem.*in January 2014

Department of Chemistry, Lakehead University and Thunder Bay Regional Research Institute,

Thunder Bay, Ontario P7B 5E1, Canada

4.1 PREFACE

In this study we identified two new fluorescent recognition probes (6-AFLU and APM-BDP) for BoNTA using computer-assisted drug design. The binding of 6-AFLU to BoNTA HC and APM-BDP to BoNTA LC enable their simultaneous use in a multi-probes fluorescence-based assay. The specific binding affinity of the identified fluorescent recognition probes were identified using FP assay. The previous identified recognition probe (PAC-BDP) exhibited binding interaction to both BoNTA HC and LC. Testing both PAC-BDP and APM-BDP against BoNTA LC using FP indicates that their simultaneous use in a multi-probe assay is feasible.

4.2 ABSTRACT

Botulinum neurotoxin (BoNTA) is one of the most dangerous toxins. The oral lethality dose of BoNTA is estimated to be 10 μg/kg for humans. To be effective, treatment of botulism must start as soon as possible after exposure. Current clinical confirmation of botulism utilizes the mouse lethality assay, which is laborious and time-consuming. The development of a rapid, simple and accurate assay to detect the botulinum neurotoxin in a variety of samples may facilitate diagnostic confirmation, reducing the starting time of botulism treatment, and may be useful in food safety preventing exposure through contaminated food products. In this study, computational and fluorescence polarization (FP) techniques were employed to successfully identify two new fluorescent recognition agents for BoNTA. Along with a previously identified fluorescent recognition agent, these compounds enable the development of a multi-probe detection assay for BoNTA. Our experimental results indicate that these three ligands bind to spatially distinct regions within the whole BoNTA structure, supporting simultaneous use. The development of a multi-probe FP-based assay to identify BoNTA enables detection of the toxin before its activation, and allows for better selectivity which reduces false positive incidence. In

addition, the FP format is suitable for high throughput implementation, enabling large-scale sample processing and may also facilitate the identification of novel inhibitors through biochemical screening of chemical libraries.

4.3 INTRODUCTION

Botulinum neurotoxin (BoNT) is produced by a gram positive bacterium, *clostridium botulinum*. There are seven serotypes of BoNT: A, B, C, D, E, F, and G[1]. The botulinum neurotoxin serotype A (BoNTA) is the most potent serotype and the causative agent of botulism, a severe neurological paralysis that affects humans. BoNT is also a potential bioterrorism weapon[1-2]. In recent years, therapeutic applications of BoNT became very attractive[3]. BoNT has been utilized for treatment of pain and a variety of neuromuscular and skeletal disabilities[4]. BoNTA is secreted as a single polypeptide molecule with the molecular weight of 150 kDa[5]. BoNTA consists of three functional domains: catalytic, ligand binding and translocation domains. The catalytic domain belongs to BoNTA's light chain (BoNTA LC), whereas the translocation and the ligand binding domains are part of the BoNTA heavy chain (BoNTA HC). BoNTs enter the human body by ingestion or inhalation of contaminated food and aerosol, or through wounds. BoNTA is classified as a zinc endopeptidase protein and the active site of the protein is located in BoNTA LC (catalytic domain). BoNT binds to nerve cells through its ligand binding domain. Upon binding to the cell surface, the binding domain transfers the catalytic domain to the cytoplasm of nerve cell[2]. The catalytic domain (BoNTA LC) with the zinc endopeptidase activity cleaves SNARE (soluble NSF-attachment protein receptor) proteins. As consequence, the dysfunctional SNARE proteins are not capable of completing the exocytosis process. The disruption of cell exocytosis process consequently blocks the secretion of acetyl choline at the neuromuscular junction, finally leading to paralysis[6].

Equine Antitoxin and BoNT vaccine are the only treatments available for botulism[7]. However these treatments are only effective before the entrance of BoNT into the cytoplasm of nerve cells[7]. Therefore, early diagnostic is essential for effective treatment of botulism. Several *in-vitro* techniques have been developed for BoNT detection including cell-based assays, PCR, immuno-PCR[8], lateral flow assay[9], ELISA and mouse lethality assay (MLA). Currently, MLA is the only accurate assay to detect all the serotypes of BoNT. However, this technique is expensive (required an animal facility and specialized staff), complex, time consuming, and ethically questionable[2, 10]. The sensitivity of MLA was reported to be 10 pg/ml[5]. Hence, developing a fast, sensitive, simple, and low-cost assay to detect BoNT is very desirable. This assay should also be suitable for high throughput screening to allow for large-scale processing of samples and to enable biochemical screening for novel therapeutic agents. Also desirable is an assay format that is suitable for portable devices. Identification of such an assay has practical applications in food safety, botulism diagnostic, defense, and research involving the medical use of BoNTA.

We chose fluorescence polarization (FP) as assay platform. FP is a powerful technique with applications in biophysical, drug discovery, analytical and clinical chemistry. Additionally, FP assay has been used for exploring protein-protein, protein-DNA and protein-ligand interactions[11]. The high sensitivity, flexibility and homogenous format of the FP method makes it an efficient technique to study protein-ligand interactions. FP is a fast and relatively cheap technique compare to MLA, and it is suitable for high throughput screening[11b, 12]. The principle behind FP is the Brownian motion theory which states that the rotation of molecules in solution will depend only on the size of molecules if the temperature and viscosity of the solvent remain constant [13]. Thus, large molecules rotate slower than the smaller molecules in solution. During an FP assay the fluorescent molecule is excited with polarized light. The emitted fluorescence is

measured in both vertical and horizontal directions to determine its degree of polarization. If the small fluorescent molecule is bound to the large protein, the emission will be more polarized compared to the free fluorescent molecule[11b, 14]. Several fluorescence based studies have been developed based on the protease activity of BoNTA. Due the fact that catalytic site of BoNTA is conserved through all the serotype, the lack of selectivity is the main disadvantage of relying on the catalytic activity of BoNTA. Additionally, fluorescent assays that depend on the toxin protease activity are not suitable to detect BoNTA in the early stage of contamination, when both HC and LC are still attached and the catalytic site in the LC is partially occluded by the HC.

In this study, we focused on other available binding pockets within the BoNTA structure for identification of novel fluorescent recognition agents[15]. We used virtual ligand screening (VLS) to identify recognition agent candidates for BoNTA detection. This approach enabled us to screen a chemical database of 1,153 compounds rapidly to identify a set of compounds which are predicted to bind against spatially distinct sites within the target protein. The initial library contained primary amines suitable for fluorescence labeling with amine reactive dyes. The virtual screening of this library identified aspartame (APM) and 6-aminofluorescein (6-AFLU) as potential recognition agents. The primary amine library was computationally labeled with a bodipy dye that is commercially available in amine-reactive form. The virtual screening of this fluorescence-labeled library indicated that a bodipy (BDP) conjugate of APM would also bind to the target. The conjugate was synthesized and tested experimentally for binding to BoNTA LC and BoNTA HC separately. 6-AFLU, an intrinsically fluorescent compound, was also tested for binding to both BoNTA LC and BoNTA HC. Our results confirmed the binding of APM, 6-AFLU and APM-BDP to BoNTA. Moreover, our results suggest that APM-BDP and 6-AFLU bind to distinct locations within BoNTA. These 3 fluorescent compounds, along with the

fluorescent recognition agent we previously identified [16], enable the development of a FP-based, multi-probe assay for BoNTA detection.

4.4 METHODS AND PROCEDURES

4.4.1 Overall Strategy

A library of 1,153 primary amines (Primary Amine Database or PAD) was screened computationally using structure-based and ligand-based pharmacophores in order to identify recognition agents targeting the catalytic site of BoNTA. PAD was also computationally docked against BoNTA using the Protein Scanning with Virtual Ligand Screening (PSVLS) approach [16]. The ligand-protein docked structures from PSVLS were scored and ranked using the original force field scores and the holistic binding scores [6]. Compounds with scores lower than a threshold of one standard deviation below the mean score for the entire library were predicted to bind to BoNTA. Among those, aspartame (APM) and 6-aminofluorescein (6-AFLU) were selected as potential recognition agents and were tested experimentally. Additional criteria for selection of hits for experimental testing is included availability, toxicity, stability and price of the compound. Selected hits from this primary screen were computationally conjugated to a bodipy dye and the conjugates were docked to BoNTA. Selected ligands from this secondary screen were synthesized by conjugation of the base primary amine to an amine-reactive dye prior to experimental testing, except hits that were intrinsically fluorescent. Binding to BoNTA was tested using fluorescence polarization. A flow chart for this strategy is presented in Figure 4.1.

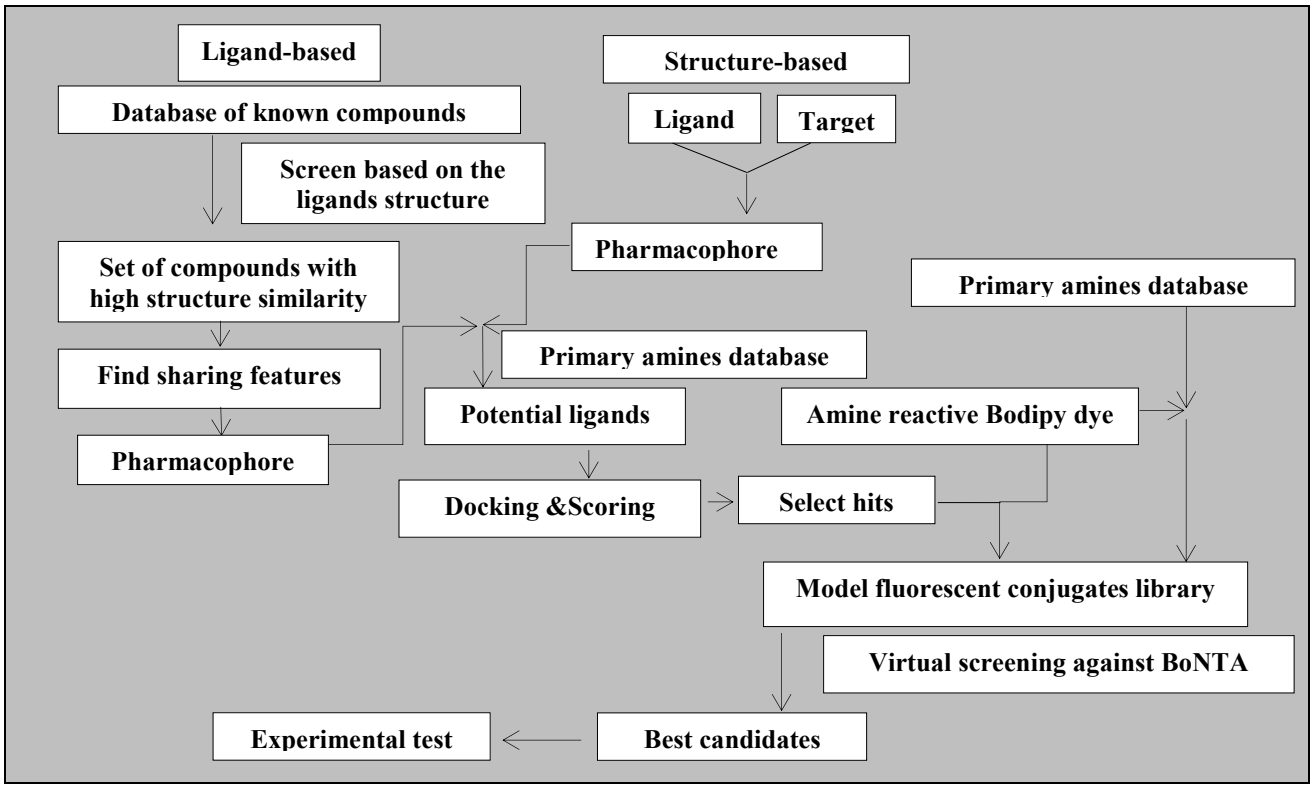


Figure 4.1 Overall strategy adopted in this study.

4.4.2 Computational Study

4.4.2.1 Identification of regions with selectivity to BoNTA

In order to identify probes with selectivity for BoNT serotype A, it is necessary to identify regions in BoNTA that differ from the other BoNT subtypes. To achieve this aim, variable residues were identified in a BoNTA protein sequence alignment. The amino acid sequences of all seven BoNT serotypes were obtained from the "UniportKB" database of the ExPASy (Export Protein Analysis System) proteomics server website. A multiple sequence alignment was constructed using Clustal X [17] with the BLOSUM30 substitution matrix, a gap opening penalty of 10, and a gap extension penalty of 0.10. The BoNT serotypes multiple sequence alignment was used to identify amino acids which are varied in BoNTA (critical residues) while conserved within the six other serotypes (serotype B, C, D, E, F and G). Special attention was given to non-

conserved (for example, polar versus non-polar) amino acids between BoNTA and the other serotypes. Additionally, the amino acids located within 7-10 Å from the center of all thirty three potential binding sites in the 3D structure of BoNTA (2NYY) were identified and compared to the critical residues from the BoNT serotypes alignment. The regions with higher number of critical residues were considered as target binding sites to differentiate BoNTA from other serotypes. A total of nine out of the thirty three regions were identified in BoNTA [16] and were prioritized for targeting: 16, 23, 8, 30, 13, 3, 19, 21 and 22.

4.4.2.2 Ligand-based and structure-based pharmacophore modeling and virtual screening

A set of 44 compounds (training set) known to bind to BoNTA [16] was used to generate a ligand-based pharmacophore model. The ligand-based pharmacophore model was generated using Pharmacophore Query Editor implemented in MOE (www.chemcomp.com). Force field (MMFF94X) types and atomic charges were assigned for each molecule. The atomic charges corresponded to the force field standard charge values. The simulations were run in gas phase with a dielectric constant of 1. The total energy was minimized for each molecule in the training set. The parameters were as follows: Hamiltonian: Force field, and Gradient: 0.05. Molecular dynamics was carried under the following parameters: Ensemble: NVT (the temperature, volume and the number of particles are constant), Duration: 100 ps, the temperature in Kelvin (T): 300, and time setup: 0.0005 ps. Flexible alignment was run after energy minimization to identify shared chemical groups between these forty four ligands (training set) and the obtained information was applied to generate a pharmacophore using MOE. The iteration limit parameter was set to 100 for the flexible alignment. The resulting of alignment had strain energy (U) of 97.47 kcal/mol, total mutual similarity score (F) of -266.32 kcal/mol and value of grand

alignment score (S) of -168.85 kcal/mol. The pharmacophore query editor (PQE) was used to generate the pharmacophore under the Polar-Charged-Hydrophobic (PCH) scheme.

A structure-based pharmacophore was also generated using three protein-ligand structures of BoNTA generated by a previous molecular docking study[16]. These protein-ligand structures were 2NYY-Paclitaxel BODIPY, 2NYY- Paclitaxel and 2NYY- Monensin. 2NYY is the identification code for the BoNTA structure used in that study, which was obtained from the Protein Data Bank (www.rcsb.org). The three 2NYY-ligand structures were superposed to create the pharmacophore model using the PQE in MOE (www.chemcomp.com) and choosing PCH as the scheme for pharmacophore generations. Both pharmacophores were applied to screen the PAD.

4.4.2.3 Molecular docking of unlabeled and fluorescence labeled libraries

The primary amines in the PAD library were computationally conjugated to BODIPY® FL-X (3-Bodipy-propanoylaminocaproic acid, N-hydroxysuccinimide ester) to form a fluorescence-labeled PAD (FL-PAD). Each labeled compound was energy minimized using the SCF (self-consistent-field) approximation in MOE. SCF is quantum mechanical approach that can be used to optimize structures of small molecules. Both PAD and FL-PAD were computationally screened against the complete (HC + LC) BoNTA structure using the same procedures described in our previous study [16]. All calculations were performed using high performance computer clusters available through SHARCNET (www.sharcnet.ca). The molecular docking parameters were the same for both libraries, except the buried surface cut-off which was lowered from 70% to 30% for the fluorescence-labeled library to compensate for the larger size of the conjugates.

4.4.3 Experimental study

4.4.3.1 Synthesis of aspartame-bodipy (APM-BDP)

To a solution of 3-bodipypropanoylaminocaproic acid *N*-hydroxysuccinimide ester (5 mg, 9.90 μmol) which was purchased from US Biological life sciences and aspartame (2.9 mg, 9.90 μmol) which was purchased from LKT Laboratories Inc in dimethylformamide (DMF) (0.5 mL) was added triethylamine (1.40 μL, 9.90 μmol) at 0 ºC. The reaction mixture was stirred at 0 ºC for 30 min and at room temperature for overnight. After reaction, DMF was removed by freeze dryer and the crude product was purified by silica gel column chromatography using methanol: dichloromethane (1:5 v/v) as the eluent to get the pure product in quantitative yield.

4.4.3.2 Photometric properties of the fluorescent compounds

Absorbance, emission and excitation spectra of 6-Aaminofluorescein (6-AFLU) and APM-BDP were obtained in the separate experiments using a Synergy 4 microplate reader. The 6-AFLU was supplied from Sigma-Aldrich. A clear microplate (greiner) was used to determine absorbance spectra and black microplate (Nunc) was utilized for emission and excitation spectra. The absorbance spectra of 6-AFLU and APM-BDP were determined in the range of 300-700 nm. The maximum absorbance from these spectra was used as excitation wavelength to obtain emission spectra. Maximum emissions were determined and used to obtain excitation spectra. A xenon lamp was used to obtain the emission and excitation spectra of the fluorescent compounds. The sensitivity of signal was set at 70 and 50 for 6-AFLU and APM-BDP, respectively.

4.4.3.3 Binding of fluorescent ligands to BoNTA HC and BoNTA LC

Fluorescence polarization (FP) was employed to explore the binding interactions of two selected fluorescent ligands to BoNTA HC and LC in separate experiments. All the FP experiments employed black Nunc 96 well microplates and a Synergy 4 microplate reader. The

excitation and emission filters for these experiments were 485/20 nm and 528/20 nm, respectively. A 510 nm full size dichroic mirror from Biotek was used in these FP assays. The excitation range of the 510 nm mirror was 440-505 nm and the emission range was 515-640 nm. A tungsten lamp was used for all the FP assays. BoNTA HC and BoNTA LC were purchased from List Biological Laboratories Inc. and reconstituted using hydrolysis buffer (20 mM HEPES buffer PH: 8.0, 1% Tween 20). The final concentration of BoNTA (either HC or LC) was fixed at 2 nM and the concentrations of fluorescent compounds were varied from 20 pM to 30 μM range in these experiments. The fluorescent compounds were dissolved in organic solvents and then diluted using assay buffer (20 mM HEPES buffer PH: 8.0, 0.1% Tween 20, 0.3 mM $ZnCl_2$) to make different concentrations. BoNTA (HC or LC) and assay buffer were added to microplate and incubated at 37 °C for 30 minutes prior to adding the fluorescent ligand to each sample. The plate was protected from light after addition of fluorescent compounds. The samples were run in duplicate and for each sample there was a duplicate control to measure the background signals. All the data was reported after deducting the background from the samples. The controls were exactly similar to the samples except that no BoNTA was present. The plate was read every 15 minutes for at least 2 hours and the data was analyzed using GraphPad Prism 6 (www.graphpad.com). To identify the range of linearity of the fluorescent compounds, a standard curve was run for each of them separately.

4.4.3.4 FP competition assay against 6-AFLU binding to BoNTA HC

In all FP competition experiments, the competitors, assay buffer and fluorescent ligand were added to the microplate and read at 37 °C as a reference prior to addition of BoNTA to the samples. The microplate was read every 15 min at 37 °C after adding BoNTA to each sample. The samples in the FP competition assays were triplicate with a final volume of 250 μl. We

found this order of adding reagents to be less complicated, more accurate and cheaper because each sample acts as its own reference before adding BoNTA. The 510 nm mirror, 485/20 nm excitation filter and 528/20 nm emission filter were used for this FP competition assay. To demonstrate competitive binding to BoNTA HC, we tested unlabeled ligands against 6-AFLU. The unlabeled ligands APM, Aminopterin (AMN) and Desmosine (DES) were tested at two concentrations using FP assay. Based on our previous computational study, AMN and DES predicted to bind to BoNTA HC and APM exhibited high score for BoNTA LC in our current study. We test APM, AMN and DES against 6-AFLU for BoNTA HC to confirm the computational prediction. AMN and DES were supplied from AXXORA company and Elastin Products Company Inc., respectively. The final concentration of 6-AFLU and BoNTA HC were fixed at 0.5 µM (EC_{50}) and 2 nM, respectively. APM, DES and AMN were dissolved in assay buffer, deionized water and DMSO, respectively, and then diluted using assay buffer. The concentration of unlabeled ligands (APM, AMN or DES) was varied in each sample.

4.4.3.5 FP competition assay against APM-BDP binding to BoNTA LC

APM-BDP was used in an FP competition assay against APM, Paclitaxel (PAC) and Paclitaxel, Bodipy ® 564/570 conjugate (PAC-BDP) binding to BoNTA LC and BoNTA HC. PAC and PAC-BDP were purchased from USBiological Life Sciences and Invitrogen Life Technologies co., respectively. PAC and PAC-BDP were dissolved in DMSO and APM-BDP was reconstituted in methanol for stock solutions. Assay buffer (20 mM HEPES buffer PH: 8.0, 0.1% Tween 20, 0.3 mM $ZnCl_2$) was used for making all the dilutions. The samples were run in triplicate at the final volume of 250 µl. The final concentrations of APM-BDP and BoNTA LC were constant at, respectively, 25 nM and 2 nM in all the samples. The 510 nm mirror with 485/20 nm excitation and 528/20 nm emission filters were used in this assay.

4.4.3.6 FP competition assay with both APM-BDP and PAC-BDP against BoNTA LC and HC

In our previous study, we were able to confirm the binding of PAC-BDP to BoNTA complex [16], however we never tested PAC-BDP against the individual LC and HC chains of BoNTA. In this FP competition assay, the concentrations of APM-BDP and BoNTA LC or HC were fixed at 25 nM and 2 nM, respectively, and the concentration of PAC-BDP was varied in different samples. Based on the maximum excitation and emission wavelengths for APM-BDP (505/515 nm) and PAC-BDP (564/570 nm), we selected two dichroic mirrors and filters combinations: 510 mirror with 485/20 nm excitation and 528/20 nm emission filters, and 570 mirror with 540/25 nm excitation and 590/35 nm emission filters. According to the Synergy 4 microplate reader manual, the excitation and emission range of Biotek 510 mirror is 440-505 nm and 515-640 nm, respectively while the excitation and emission range of 570 mirror is reported to be 515-565 and 575-735 nm, respectively (Figure 4.2.A and B). We employed both mirrors to determine the binding interactions of APM-BDP and PAC-BDP against BoNTA LC and HC.

A) B)

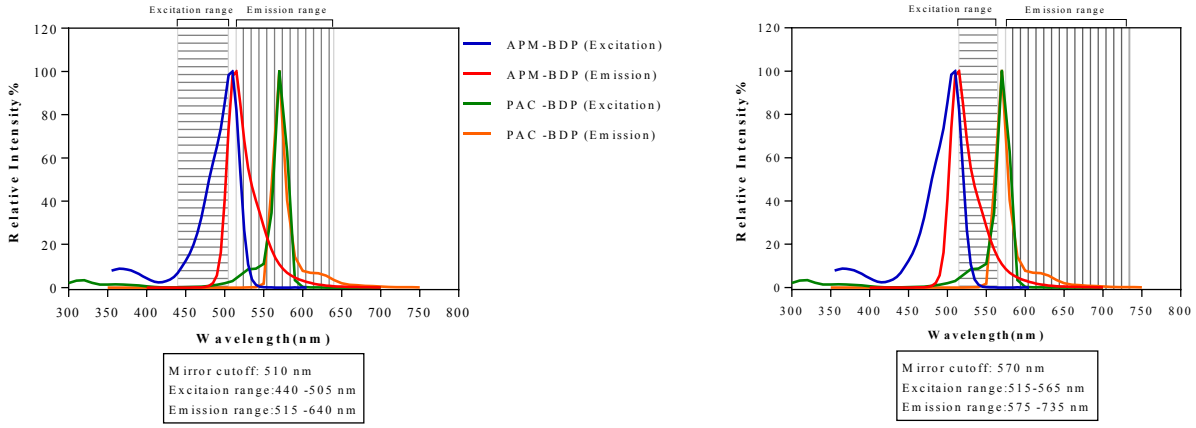


Figure 4.2 Excitation and emission spectrum of APM-BDP and PAC-BDP are shown in both graphs (A and B). In graph A, the parallel lines mark the excitation and emission wavelength ranges of the 510 nm dichroic mirror. In graph B, the parallel lines represent the excitation and emission ranges of the 570 nm mirror. The blue and red lines represent the excitation and emission spectrum of APM-BDP. The green and orange lines correspond to the excitation and emission spectrum of PAC-BDP, respectively.

4.5 RESULTS AND DISCUSSION

4.5.1 Pharmacophore modeling

In general, ligand-based pharmacophore modeling compares different conformations of ligands to construct a pharmacophore model that accounts for the binding of these ligands to the target protein. In this study, three common features (F1, F2, and F3) were identified between the molecules with 100%, 93% and 57% scoring percentages, respectively. Each feature represents a set of atoms with a physicochemical property that is important for binding (Figure 4.3). F1 represents ML |Acc | Don which means in order to match to this feature a ligand conformation needs to have either a metal ligator (ML), a hydrogen bond acceptor (Acc), or a hydrogen bond donor at this position. F2 represents Aro | Don; to match this feature a ligand conformation needs to have either an aromatic feature (Aro), or a hydrogen bond donor (Don) at this position.

F3 feature represents Hyd| ML|Aro|Don which is matched by compounds having either a hydrophobic region (Hyd), a metal ligator (ML), aromatic feature (Aro), or hydrogen bond donor (Don) at this position.

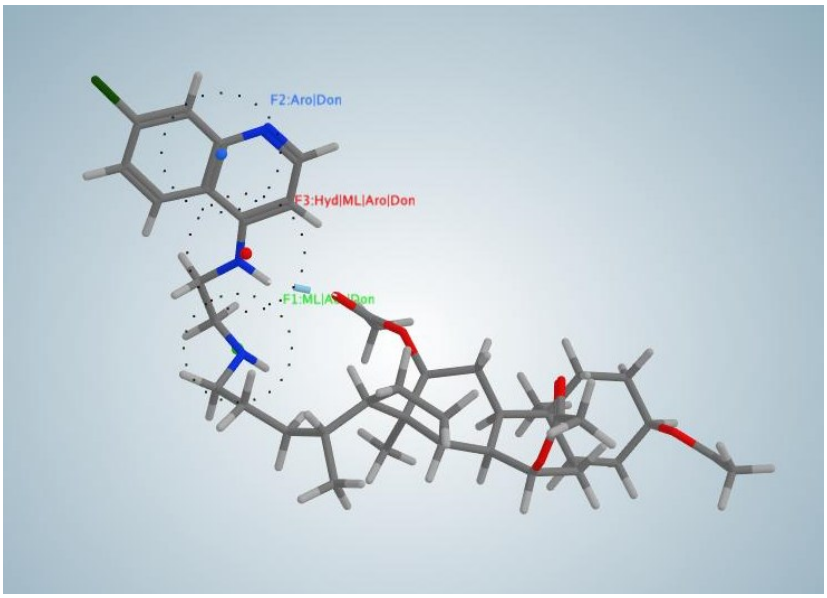


Figure 4.3 Ligand-based pharmacophore model shows the mapping of the identified features onto one of the compounds in the training set. ML: metal ligator, Acc: hydrogen bond acceptor, Aro: aromatic feature and Don Hydrogen bond donor. The pharmacophore features are shown as dotted circles.

The structure-based pharmacophore was built with three features with 100% scoring, selected out of 22 features. F1 represents Hyd which corresponds to hydrophobic character. The F2 represents ML | Hyd | Don; in order to match this feature a ligand needs to have either a metal ligator (ML), a hydrophobic moiety or a hydrogen bond donor (Don) at this position. F3 represents ML |Hyd | Acc | Don which means in order to match this feature a ligand needs to have either a metal ligator (ML), a hydrophobic moiety(Hyd), a hydrogen bond acceptor (Acc), or a hydrogen bond donor (Don) at this position (Figure 4.4).

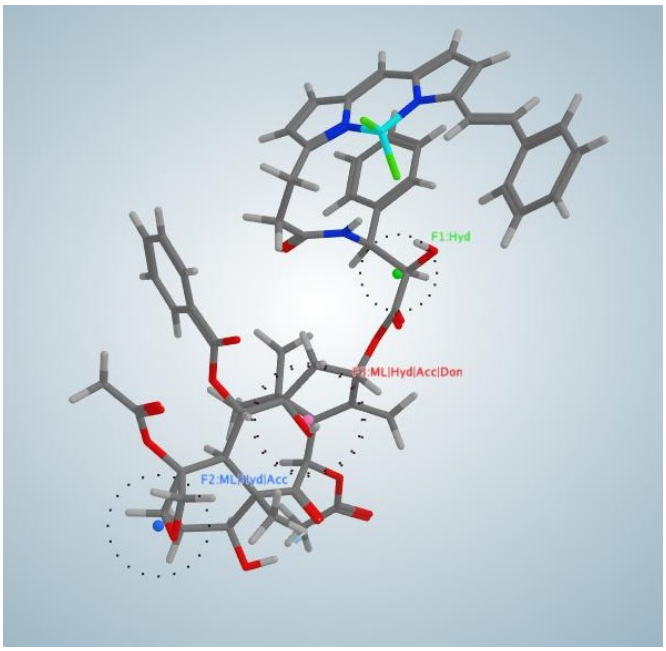


Figure 4.4 Structure-based pharmacophore model shows the mapping of the identified features onto Paclitaxel BODIPY. ML: metal ligator, Acc: hydrogen bond acceptor, Hyd: hydrophobic feature and Don hydrogen bond donor. The pharmacophore features are shown as dotted circles.

The generated pharmacophores were applied to pre-filter PAD using pharmacophore search option in MOE in order to identify molecules likely to bind to the catalytic site of BoNTA. During the pharmacophore-based virtual screening the pharmacophore model was matched to different conformations of the molecules in the library. The best conformation with the lowest RMSD (Root Means Square Deviation) of each molecule was selected. The number of selected hits after the screening of PAD using the ligand-based and structure-based pharmacophores were 1,100 and 303, respectively, out of 1,153 compounds comprising the PAD. Since both ligand-based and structure-based pharmacophores were constructed based on ligands that bind to the same binding site (active site), compounds that match these pharmacophores are expected to bind to the same site of BoNTA. The number of surviving compounds from the ligand-based pharmacophore screening indicates that this pharmacophore was not discriminative, as most of the compounds in the library survived the screening. In

contrast, the structure-based approach reduced the library to 303 candidates. APM, a compound that also survived the primary screening of the PAD library (see section "4.5.2"), is among these compounds. A second compound, 6-AFLU, predicted to bind to BoNTA HC in the primary screening of the PAD library, did not pass the structure-based pharmacophore screening. These results are in agreement with experimental results discussed in sections 4.5.9 and 4.5.8 for APM and 6-AFLU. This agreement indicates that structures predicted by PSVLS, which were used for pharmacophore modeling, are fair representations of the key interactions between protein and ligand.

4.5.2 Virtual screening of unlabeled ligands

The results from the virtual screening of the PAD library were processed using three scoring schemes: the original force field-based scores reported by HierVLS[18], holistic scores [16], and Principal Components Analysis (PCA). These scores are provided as supporting documents. The force field scores and the holistic scores of the screened ligands were plotted individually versus the number of ligands for each of the nine regions predicted to show selectivity for BoNTA (Appendix A). The ligands were ranked based on their scores (force field or holistic), and ligands with scores lower than one standard deviation of the mean for the entire library were considered to be potential binders of BoNTA experimentally. The ligands were ranked according to their force field and holistic scores at each particular binding region. The ligands were required to pass the threshold for at least one of the scoring schemes in order to be considered as potential binder. For both force field and holistic scores lower (more negative) numbers means better binding affinity. As discussed before, we are interested on ligands passing the binder/non-binder threshold at the nine binding sites with potential selectivity to BoNTA. Therefore the surviving compounds in each region of interested were explored for selection of recognition

agent candidates. In addition to passing the binder/non-binder thresholds, candidates were selected according to parameters including commercial availability, structure, price, known toxicity, and chemical stability. Aspartame (APM) was one of the candidates selected for further analysis. This compound, which is a known artificial sweetener, had the proper structure for fluorescence labeling and it was readily available for purchase at research-grade purity. The other candidate selected for testing was 6-Aminofluorescein (6-AFLU) which is an intrinsically fluorescent molecule and, thus, was ready for fluorescent polarization experiments. Based on the holistic and force field scores, APM passes the binder/ non-binder threshold only in regions 16 and 21. APM binds best in region 21, followed by region 16, both located in the catalytic domain of BoNTA. The score versus ligand graph for APM in regions 21 and 16 are shown in Figure 4.5.

Based on the force field scores, 6-AFLU passed the binding thresholds in five of the 9 binding sites considered, best scoring in region 8 which is located in the binding domain of BoNTA. 6-AFLU scored lower than the threshold in regions 23, 13, 19, 21 and 8 based on force field scores, although their holistic scores were not lower than threshold (but very close to threshold) for all these regions. The holistic and force field scores for the entire screening library docked to region 8 are provided in Figure 4.6.

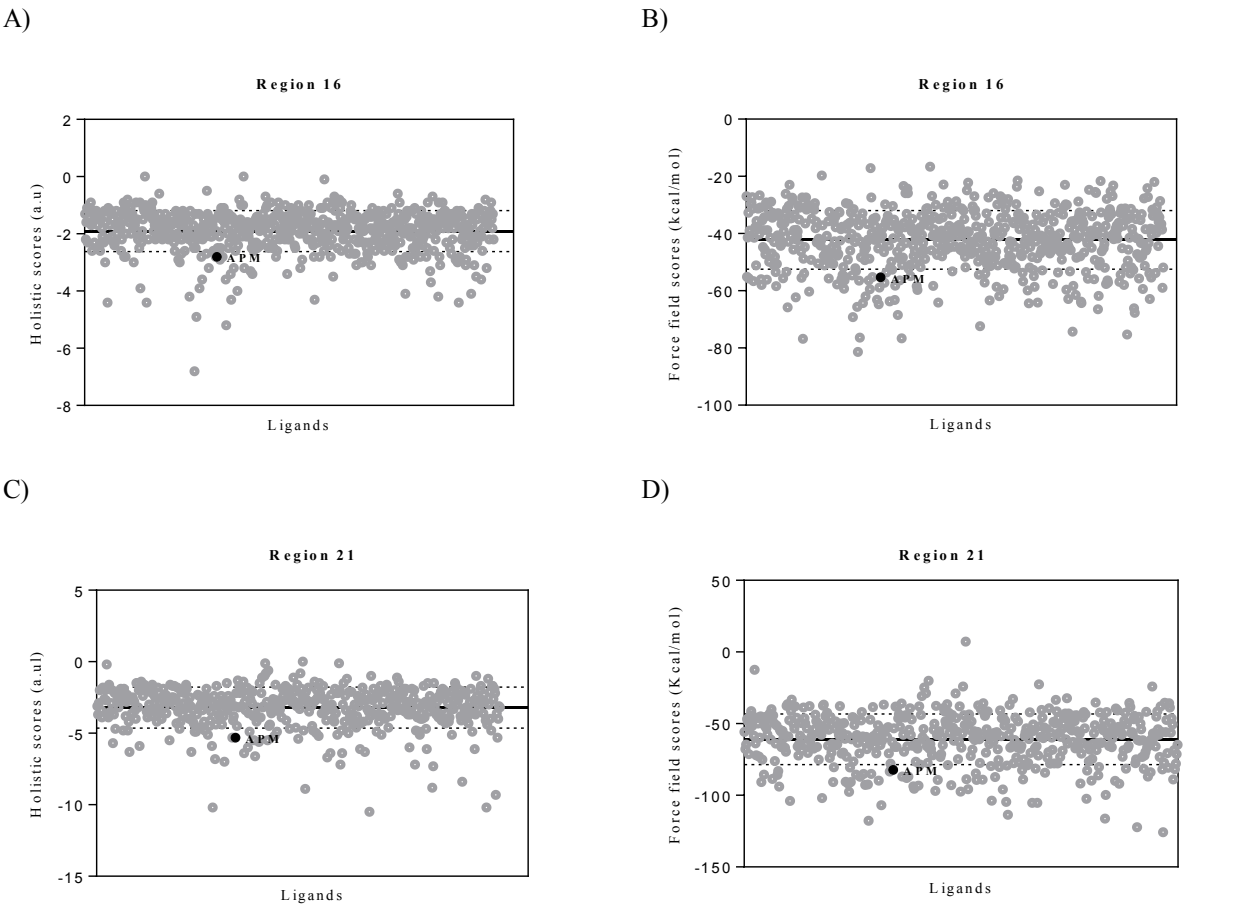


Figure 4.5 (A) holistic and (B) force field binding scores of all the ligands in the virtual library docked to region 16 in the BoNTA structure. (C) holistic and (D) force field binding scores of all the ligands in the virtual library docked to region 21 in the BoNTA structure. The solid lines in the graphs represent the average of binding scores of all ligands. The dotted lines mark one standard deviation above or below the average. The lower dotted line corresponds to the non-binder/binder threshold. APM exhibited scores lower than one standard deviation in both scoring schemes and thus passed the binder/non-binder threshold.

A) B)

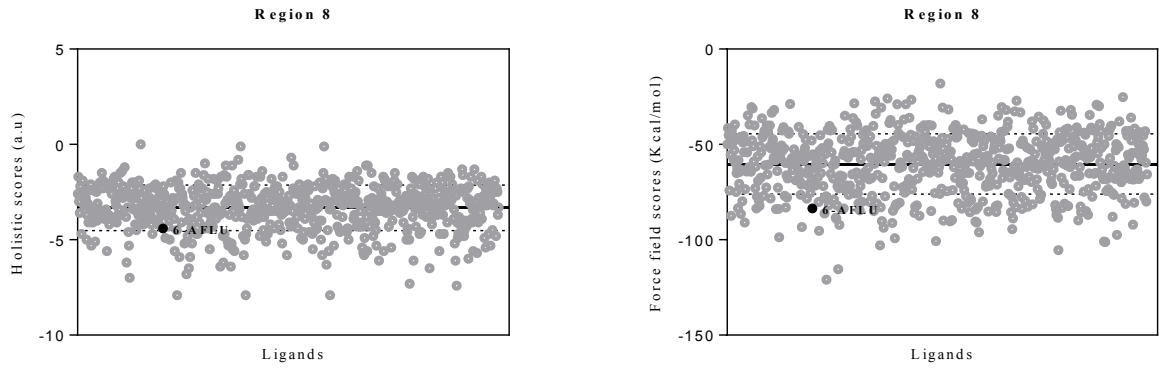


Figure 4.6 (A) holistic and (B) force field binding scores of all the ligands in the virtual library docked to region 8 in the BoNTA structure. The solid lines in the graphs represent the average of binding scores of all ligands. The dotted lines mark one standard deviation above or below the average. The lower dotted line corresponds to the non-binder/binder threshold. 6- AFLU exhibited scores lower than one standard deviation Based on the original binding score graph (B) however 6-AFLU holistic scores was on the border of the binder/non-binder threshold (A).

4.5.3 Virtual screening of fluorescence labeled ligands

The molecular docking result for the fluorescence labeled library was analyzed using the same strategy described for the unlabeled ligands. During molecular docking, scores (force field and holistic) were calculated for each fluorescent ligand bound to each individual binding region within the structure of BoNTA. The fluorescent ligands were ranked based on each scoring scheme (force field or holistic) in the 9 binding regions of interest (selective for BoNTA) (Appendix B). According to both force-field and holistic scores, APM-BDP survived the binder/non-binder threshold in regions 21 and 23 which are located in BoNTA LC. The holistic score for APM-BDP was on the threshold line in region 21, however APM-BDP scores were lower than one standard deviation in region 23 (Figure 4.7) for both force field and holistic scores. APM thus was found to bind well in region 21 before and after fluorescence labeling. In contrast, fluorescence labeling of APM affected its affinity for region 16, and APM-BDP did not

survive the threshold in this region. Interestingly, APM did not pass the threshold in region 23 in the first molecular docking but, after labeling, the structure of APM-BDP became more suitable to bind to region 23 (based on both scoring schemes), with binding scores that indicate higher affinity to region 23 (-87.2) than to region 21 (-74.9).

A) B)

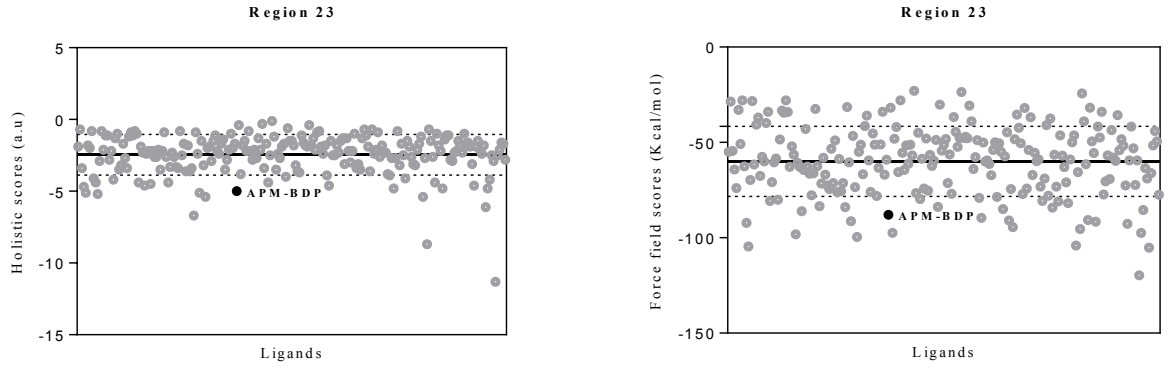


Figure 4.7 (A) holistic and (B) force field binding scores of all the ligands in the virtual library docked to region 23 in the BoNTA structure. The solid lines in the graphs represent the average of binding scores of all ligands. The dotted lines mark one standard deviation above or below the average. The lower dotted line corresponds to the non-binder/binder threshold. APM-BDP exhibited scores lower than one standard deviation in both scoring schemes and thus passed the binder/non-binder threshold.

4.5.4 Principal components analysis (PCA) of molecular docking results

A third scoring scheme was used to select probe candidates for experimental testing. Principal Component Analysis was used to combined the original force field scores for each ligand docked to each of the 33 binding region in BoNTA into a orthogonal set of scores (principal components), each accounting for some variance of the original data. Generally, the first component (PC1) accounts for most of the variance and, for our purpose, is a score representative of the overall binding affinity of the ligand for the target protein. The PAST

statistical program (version 2.17) was used to carry out this analysis [19]. PCA was used to transform the molecular docking results into first principal component scores for both unlabeled and fluorescence-labeled ligands, as well as the combination of fluorescent labeled and unlabeled libraries. The attempt to use PCA was unsuccessful because of the low variance carried by the first component (% 47 and % 53.1). The PCA analysis graphs and the table of % variance provided in appendix C.

4.5.5 Absorbance and fluorescence spectra for 6-AFLU

6-AFLU was dissolved in acetone and then diluted using assay buffer. The concentration of acetone in the samples was less than 2%. The final concentration of 6-AFLU was at 20 µM in the samples. The absorbance spectrum of 6-AFLU is shown in Figure 4.8.A. Based on this spectrum, the maximum absorbance was determined to be 490 nm. The maximum absorbance (490 nm) was used as the excitation wavelength to obtain the emission spectrum (Figure 4.8.B), and the maximum emission obtained from this spectrum was applied to determine the excitation spectrum of 6-AFLU (Figure 4.8.B). According to these spectra, the maximum excitation and emission wavelengths for 6-AFLU are 490 nm and 510 nm, respectively.

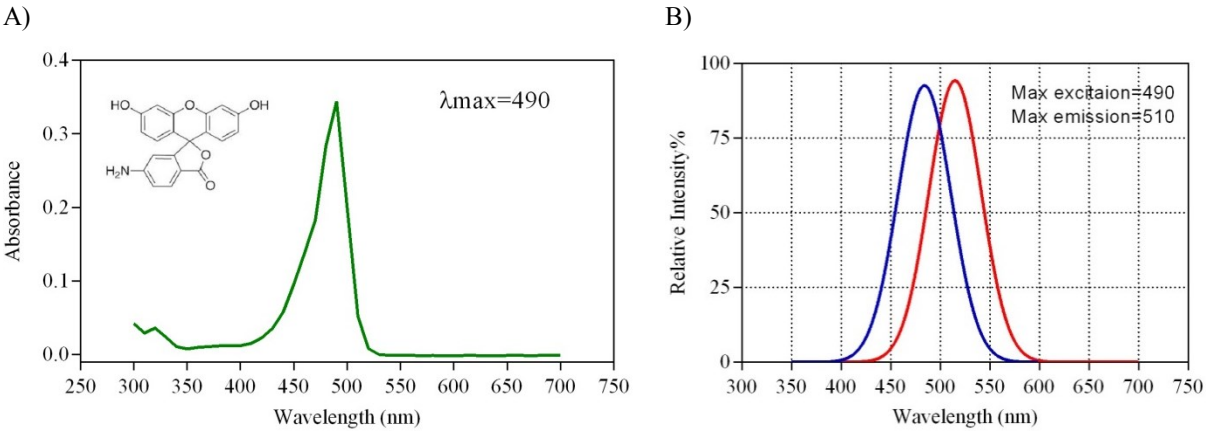


Figure 4.8 Absorbance spectrum (A) and excitation and emission spectra (B) for 6-AFLU.

4.5.6 Absorbance and fluorescence spectra for APM-BDP

The absorbance, excitation and emission spectrum of APM-BDP was determined using the same procedure described for 6-AFLU. Methanol was used as a solvent for APM-BDP and the samples were diluted using assay buffer. The concentration of methanol in the samples was less than 2%. The final concentration of APM-BDP was 20 µM. The maximum absorbance of APM-BDP was obtained at 505 nm according to the spectrum presented in Figure 4.9.A. In addition, the excitation and emission spectra for APM-BDP, displayed in Figure 4.9.B, yield maximum excitation and emission wavelengths of 505 nm and 513 nm, respectively.

A) B)

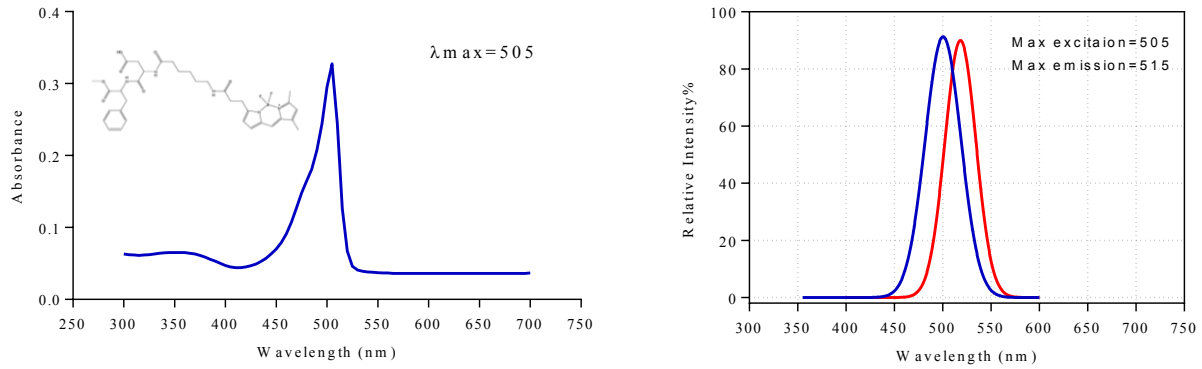


Figure 4.9 Absorbance spectrum (A) and excitation and emission spectra (B) for APM-BDP.

4.5.7 Analysis of the LC/MS and NMR data of APM-BDP

The NMR and LC-MS analytical methods were used for verification and characterization of APM-BDP (Appendix D). ^{1}H NMR (500 MHz, CD_3OD) δ 1.16-1.22 (m, 2H), 1.34-1.40 (m, 2H), 1.42-1.51 (m, 2H), 2.08 (t, J = 7.2 Hz, 2H), 2.17 (s, 3H), 2.41 (s, 3H), 2.46-2.54 (m, 3H), 2.61-2.69 (m, 1H), 2.87-2.94 (m, 1H), 2.98-3.08 (m, 3H), 3.12 (t, J = 7.2 Hz, 2H), 3.58 (s 3H), 4.55 (dd, J = 5.8, 7.7 Hz, 1H), 4.62 (dd, J = 5.8, 7.7 Hz, 1H), 6.11 (s, 1H), 6.21 (d, J = 4.2 Hz, 1H), 6.90 (d, J = 4.2 Hz, 1H), 7.04-7.19 (m, 5H), 7.32 (s, 1H). MS $(M + Na)^+$ 704.3.

4.5.8 Binding of 6-AFLU against BoNTA HC and BoNTA LC

Preliminary binding tests were run to optimize the concentration of BoNTA in the FP binding assay, and the final concentration of 2 nM was selected for further testing. A binding curve obtained by FP for 6-AFLU against BoNTA HC (Figure 4.10.A) demonstrates dose-dependent increase in FP value which results from the specific binding interactions between 6-AFLU and BoNTA HC. A three parameters non-linear regression (log (agonist) versus response) in GraphPad Prism 6 (www.graphpad.com) was used to fit the dose response curve and estimate EC50. The calculated EC50 of 6-AFLU against BoNTA HC was 546 ± 60nM. Based on our experiments, BoNTA and ligands reach equilibrium after 60 min so all the data provided in this study are for readings after 60 min.

Binding of 6-AFLU was tested using FP against BoNTA LC at four concentrations (100 pM, 100 nM, 1 μM and 50 μM) as shown in Figure 4.10.B. However, no significant difference between samples with BoNTA LC and samples without BoNTA LC was observed. Additionally the FP was dependent on the concentration of 6-AFLU, with FP decreasing due to increase in the number of unbound fluorescent molecules in each sample. According to the computational results the best score of 6-AFLU was at region 8, located in binding domain of BoNTA (BoNTA HC). Thus, the experimental results are in agreement with PSVLS prediction that 6-AFLUbinds preferentially to BoNTA HC over BoNTA LC.

A) B)

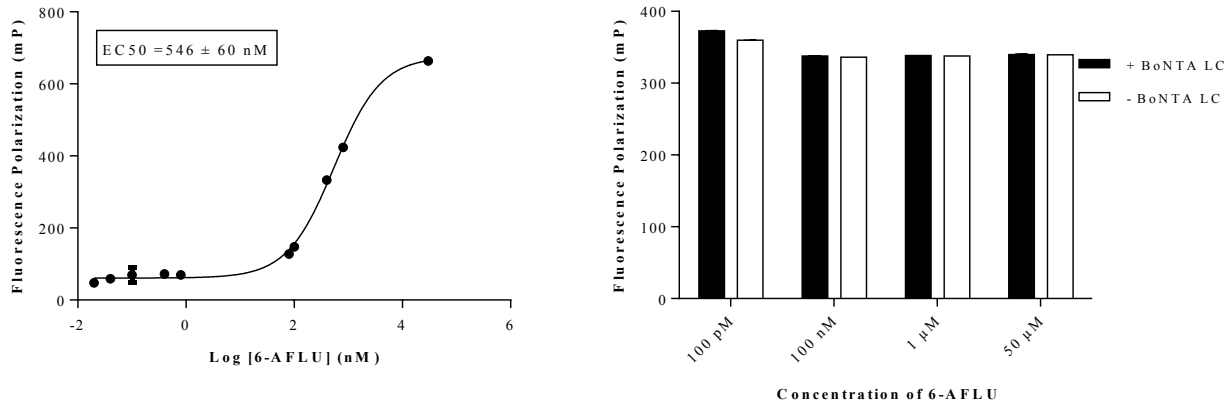


Figure 4.10 (A) 6-AFLU and BoNTA HC binding curve with an EC_{50} of 546±60nM using FP assay. (B) 6-AFLU tested for binding against BoNTA LC shows no significant interaction to BoNTA LC. The background signals were subtracted from each sample before calculating the fluorescence polarization. The graphs are based on two separate experiments. The concentration of BoNTA HC (A) and BoNTA LC (B) were fixed at 2 nM. Samples were run in duplicate in both assays. Binding assays were conducted at body temperature (37°C) with an incubation period of 30 minutes.

4.5.9 Binding curve of APM-BDP against BoNTA LC

A binding curve obtained by FP for APM-BDP against BoNTA HC (Figure 4.11.B) demonstrates dose-dependent increase in FP value which results from the specific binding interactions between APM-BDP and BoNTA LC. A three parameters non-linear regression (log (agonist) versus response in GraphPad Prism 6 (www.graphpad.com) was used to fit the dose response curve and estimate EC_{50}. The calculated EC_{50} of APM-BDP for BoNTA LC was 20.96 ± 10 nM.

Binding of APM-BDP was tested using FP against BoNTA HC at four concentrations (100 pM, 1 nM, 100 nM and 1 µM) as shown in Figure 4.11.A. However, no significant difference between samples with BoNTA HC and samples without BoNTA HC was observed (Figure 4.11.A). Additionally the FP was dependent on the concentration of APM-BDP, with FP

decreasing due to increase in the number of unbound fluorescent molecules in each sample. Thus we conclude that APM-BDP binds to BoNTA LC with an EC_{50} of 21 nM, but it does not bind to BoNTA HC at concentrations lower than 100 μM.

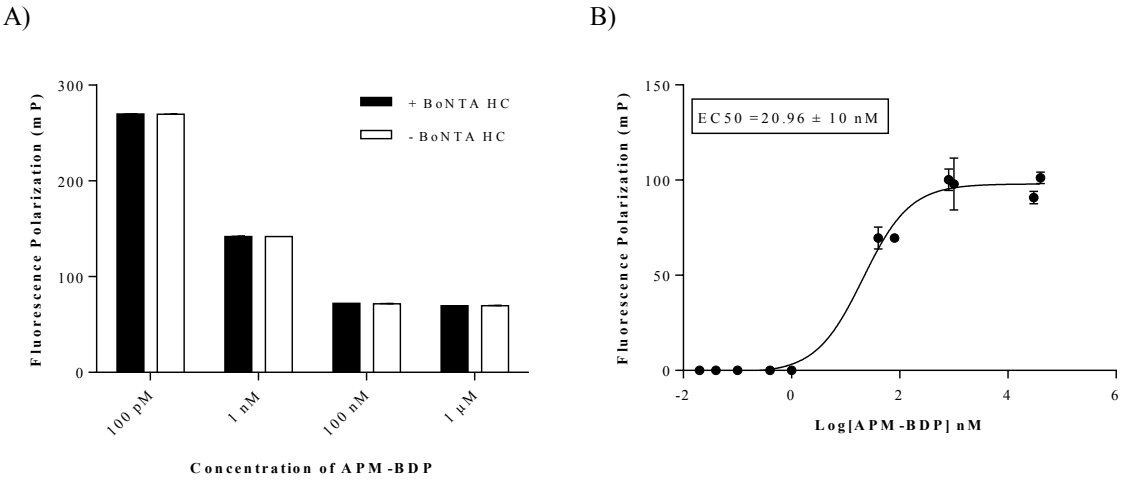


Figure 4.11 (A) APM-BDP tested for binding against BoNTA HC shows no significant interaction to BoNTA HC (B) APM-BDP and BoNTA LC binding curve with an EC_{50} of 20.96 ± 10 nM using FP assay. The background signals were subtracted from each sample before calculating the fluorescence polarization. The graphs are based on two separate experiments. The concentration of BoNTA HC (A) and BoNTA LC (B) were fixed at 2 nM. Samples were run in duplicate in both assays. Binding assays were conducted at body temperature (37 °C) with an incubation period of 30 minutes.

4.5.10 FP competition assays against 6-AFLU binding to BoNTA HC

APM, DES, and AMN were tested in competition with 6-AFLU against BoNTA HC to confirm our computational prediction of binding affinity of DES and AMN for BoNTA HC and no interaction of APM against BoNTA HC (Figure 4.12). To account for the effect of solvent in the FP measurement, a control was prepared for each individual sample. The concentration of solvent in the control was similar to its related sample. In other words, all controls contain only 6-AFLU and BoNTA HC at the same concentration than the samples and should have the same

FP (maximum signal). However there are slight differences between the controls (chess pattern bars in Figure 4.12) which are due to the concentration of different solvents. The FP values were calculated after subtracting the references (the samples were read before adding BoNTA HC). The black bar samples contained 6-AFLU, unlabeled ligand and BoNTA HC. As evident in Figure 4.12 the FP signal was not changed in the presence of APM at the concentrations of 100 and 1 μM. It means APM did not show binding interaction to BoNTA HC and no sign of competition between APM and 6-AFLU was observed. This result was predictable considering the negative result of APM-BDP binding study against BoNTA HC (Figure 4.11.A). In addition, the data showed that the FP signals were decreased in the presence of AMN (black bars) compare to the sample with only 6-AFLU and BoNTA HC (chess pattern bars). Furthermore, the FP signal reduction (sample with AMN) was dose-dependent. This outcome was explained that 6-AFLU and AMN were binding to the same binding site of the BoNTA HC. The FP signal was reduced because if the two ligands compete to the same region, in the presence of AMN some of the 6-AFLU molecules were replaced by AMN molecules therefore the number bond 6-AFLU molecules (fluorescent molecule)were reduced and subsequently, the FP was decreased by increasing the AMN concentration. In the FP competition assay of DES and 6-AFLU, The FP responses were increased in the presence of DES (black bars) compare to the sample without DES (chess pattern bars) and the higher concentration of DES was shown the higher FP signal. This result can be explained if DES and 6-AFLU bind to two different binding regions of BoNTA HC. In this case the size of the complex molecule (6-AFLU-BoNTA HC-DES) was increased and this was explained having FP signal higher than expected maximum FP signal. In the other study, we were able to identify the binding interaction of radiolabeled AMN and DES to BoNTA HC using Scintillation proximity assay (in preparation) and these data were

confirmed each other. The white bar correspond to the sample which was contained only 6-AFLU without BoNTA HC. The FP signal of this sample was considered as a baseline signal.

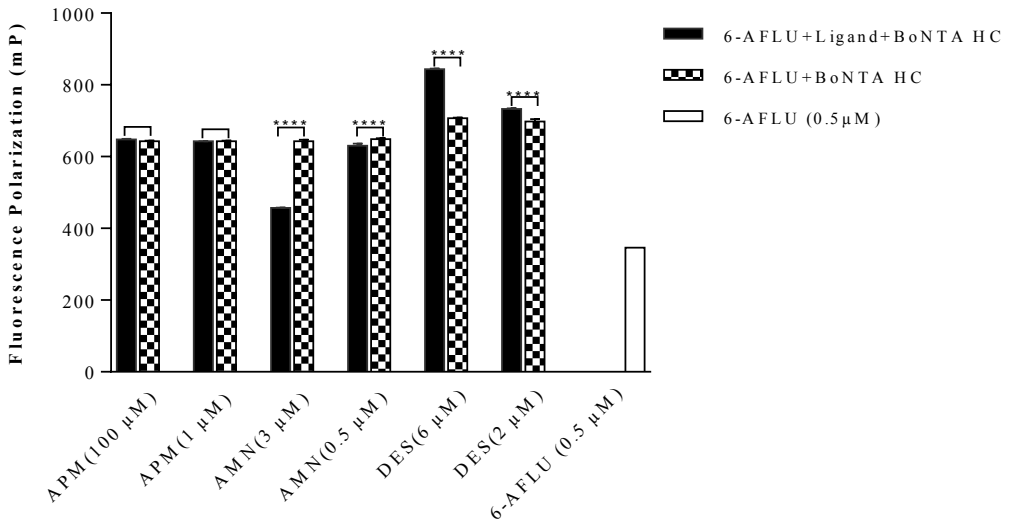


Figure 4.12 FP competition assay against 6-AFLU binding to BoNTA HC for APM, AMN, and DES BoNTA HC. Each unlabeled ligand was tested at two different concentrations. In this graph, black bars represent FP in presence of a competitor (6-AFLU + unlabeled ligand + BoNTA HC) and chess pattern bars represent the FP resulting from the fluorescent agent binding to the protein (6-AFLU + BoNTA HC). The two values were compared to each other to determine the effect of the unlabeled ligands. The white bar represents the baseline signal as it contains only 6-AFLU and no competitor or toxin. APM did not affect the FP associated to 6-AFLU binding to BoNTA HC and, therefore, we conclude that APM does not bind to BoNTA HC. AMN and DES both change the FP response. AMN decreases FP, indicating competition with 6-AFLU for the same binding site in BoNTA HC. DES increases FP, indicating that DES and 6-AFLU bind to the different binding sites in BoNTA HC. FP is dose-dependent in both cases, indicating specific binding. A 510 nm dichroic mirror with 485/20 nm excitation and 528/20 nm emission filters were used for this FP competition assay. The error bars in the graph represent standard deviations and the stars indicate significant differences of FP signal in the samples with and without BoNTA HC according to two way Anova analysis.

4.5.11 FP competition assay of APM-BDP for BoNTA LC

The Figure 4.13 displays the results of competition assays for APM-BDP, PAC and PAC-BDP against BoNTA LC. In this graph, black bars represent samples containing APM-BDP, competitor and BoNTA LC, and chess pattern bars represent samples containing APM-BDP and

BoNTA LC with no competitor. The uninhibited FP signals (chess pattern bars) for all the samples related to PAC and PAC-BDP (dissolved in DMSO) were almost the same. However FP was higher for samples related to APM which were dissolved in assay buffer (chess pattern signals). As previously discussed, the solvents used to solubilize the competitors had an effect on the ligand binding interaction and subsequently on the FP signals. The only difference between all the chess pattern bars samples is the presence of the solvents of competitor at the same concentrations of their related black bar. The white bar in this graph (Figure 4.13) shows the FP signal of APM-BDP alone at the concentration of 25 nM, which was fixed in the competition assays. By comparing the signal of each black bar to its corresponding chess pattern bar we concluded that all competitors APM, PAC and PAC-BDP bind to BoNTA LC. APM competes with APM-BDP for the same binding region of BoNTA LC as evident from the FP signal reduction in the presence of APM (competitor in black bars). PAC and PAC-BDP bind to different binding sites of BoNTA LC causing the FP signal induced by PAC-BDP to increase in the presence of the competitor PAC. In addition, the FP responses were dose-dependent in all three cases. The binding of PAC-BDP to BoNTA complex and the IC_{50} of PAC against BoNTA LC (5.2 μM) were determined in our previous paper using FP and FRET assays, respectively [16]. The computational results indicated that APM-BDP should not bind to the active site of BoNTA LC. The experimental results confirm our computational prediction, the FP signal increases in the presence of APM-BDP and PAC (inhibitor for the catalytic activity of BoNTA LC) which indicates APM-BDP and PAC binding to two different pockets of BoNTA LC (Figure 4.13).

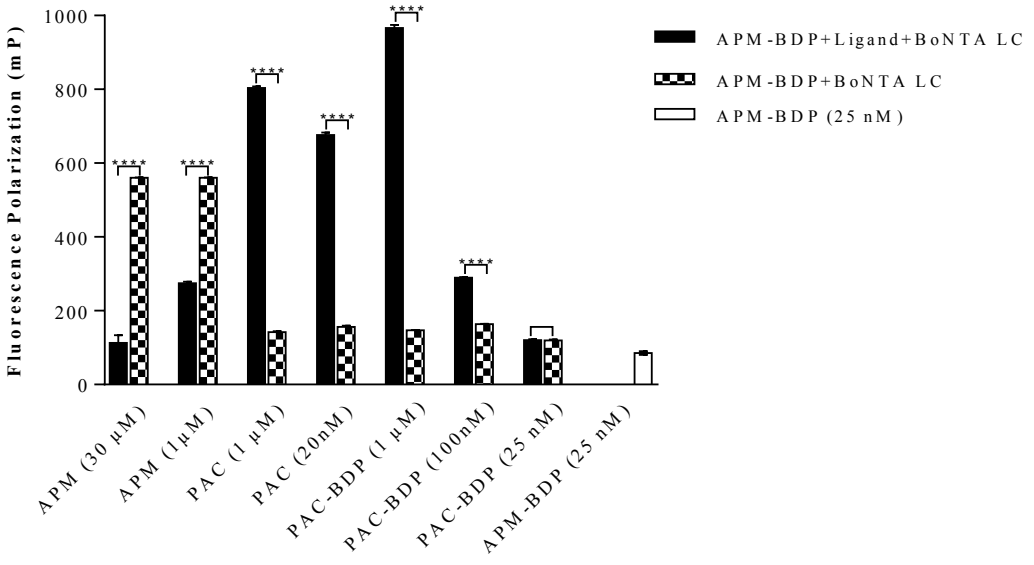


Figure 4.13 FP competition assay against APM-BDP binding to BoNTA LC for APM, PAC, and PAC-BDP. Each unlabeled ligand was tested at two or three different concentrations. In this graph, black bars (APM-BDP + Competitor + BoNTA LC) and chess pattern bars (APM-BDP + BoNTA LC) were compared to each other to determine the effect of competitors. The white bar represents the baseline signal as it contains only APM-BDP and no competitor or toxin. APM was decreased the FP associated to APM-BDP binding to BoNTA LC and, therefore, we conclude that APM binds to the same binding site of BoNTA LC. PAC and PAC-BDP both enhance the FP response, indicating that PAC and PAC-BDP bind to the different binding sites in BoNTA LC. FP dose-dependent in both cases, is indicating specific binding. A 510 nm dichroic mirror with 485/20 nm excitation and 528/20 nm emission filters were used for this FP competition assay. The error bars in the graph represent standard deviations and the stars indicate significant differences of FP signal in the samples with and without BoNTA LC according to two way Anova analysis.

4.5.12 FP competition assay of APM-BDP and PAC-BDP for BoNTA LC and HC using two mirror sets

In our previous study, we were able to confirm the binding of PAC-BDP to BoNTA complex, however we never tested PAC-BDP against BoNTA LC and HC separately. In this FP competition assay, the concentration of APM-BDP and BoNTA (LC or HC) were fixed at 25 and 2 nm, respectively, and the concentration of PAC-BDP was varied in different samples. Based on the maximum excitation and emission wavelengths of APM-BDP (505/515 nm) and PAC-BDP

(564/570 nm), two mirror and filters sets were selected to conduct FP competition assays: 510 nm mirror with 485/20 nm excitation and 528/20 nm emission filters, and 570 nm mirror with 540/25 nm excitation and 590/35 nm emission filters.

Results obtained with the 510 nm mirror set are shown in Figure 4.14. This figure exhibits competition assay results for APM-BDP and PAC-BDP against BoNTA LC (A) and BoNTA HC (B). The data indicates that the FP signal was higher in presence of both fluorescent compounds (PAC-BDP and APM-BDP) compared to the samples with only one fluorescent compound (PAC-BDP). Clearly, this effect confirms the binding of both APM-BDP and PAC-BDP to different binding sites of BoNTA LC. Moreover, since the concentration of APM-BDP was constant in all samples, the decrease in FP observed in samples with lower concentrations of PAC-BDP confirms the synergistic binding. In samples with only PAC-BDP (stripe pattern bars), FP was also concentration-dependent. The FP of the samples with only APM-BDP (at constant concentration, dotted pattern) is slightly different due to the effect of solvent. The concentration of solvent in each dotted pattern bar are exactly the same as its related black bar. Additionally, FP associated to PAC-BDP and APM-BDP without BoNTA LC was determined as baseline of the FP readings (white column). The data was normalized before plotting. In the case of BoNTA HC (Figure 4.14.A), FPs were not significantly changed for the samples with (black bars) and without APM-BDP (stripe pattern bars). This result indicates that there is no binding between APM-BDP and BoNTA HC at the concentrations tested. Additionally, FP associated to APM-BDP binding to BoNTA HC (dotted pattern bars) are lower than the baseline (white column) which confirmed that APM-BDP does not bind to BoNTA HC.

A) B)

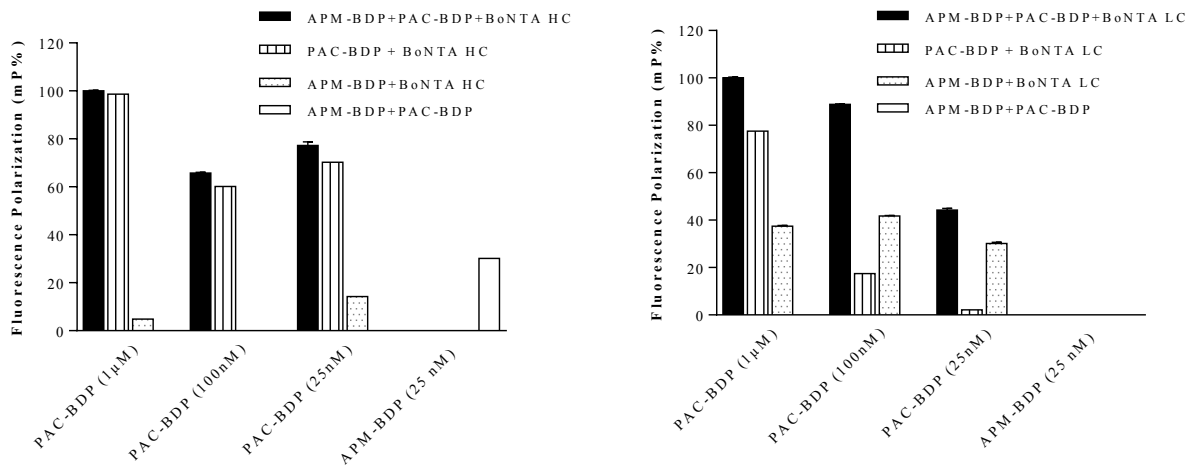


Figure 4.14 Competition assays between APM-BDP and PAC-BDP against BoNTA HC (A) and BoNTA LC (B). In both experiments, the concentrations of APM-BDP and BoNTA (HC or LC) were fixed and the concentration of PAC-BDP was varied. The black bars represent the sample with both fluorescent compounds and BoNTA. Striped pattern bars correspond to the mixture of PAC-BDP and BoNTA. The dotted pattern columns represent FP associated to APM-BDP bound to BoNTA (HC or LC). The white bar represents the baseline FP corresponding to only APM-BDP or PAC-BDP. The 510 mirror, 485/20 excitation filter and 528/20 emission filter were used for the both FP competition assays.

The FP competition assay between APM-BDP and PAC-BDP against BoNTA LC and BoNTA HC were also ran using the 570 nm mirror set (Figure 4.15) due to the lack of availability of proper set of mirror and filter to detect both fluorophores signal at the same time. Additionally the competition assay using the 570 nm mirror was ran to reconfirm the results obtained with the 510 nm mirror (Figure 4.15). The FP measurement for samples with only APM-BDP and BoNTA (HC or LC) were equal to the negative control (assay buffer). This means that using the 570 nm mirror set we were only able to measure FP from a PAC-BDP point of view. However, the size of the complex should increase if APM-BDP and PAC-BDP bind at the same time to BoNTA LC. This explains the differences between black and stripe bars (Figure 4.15.B). According to the result presented in the Figure 4.15.B, the black bars (APM-BDP

+PAC-BDP + BoNTA LC) at the concentrations of 1 μM and 25 nM show higher FP compared to their related stripe bars (PAC-BDP + BoNTA LC). This result supports that APM-BDP and PAC-BDP bind to the BoNTA LC. The graph A of Figure 4.15 shows that there are no significant differences of FP in the presence (black bars) or absence (stripe pattern bars) of APM-BDP which also confirm that APM-BDP does not bind to the BoNTA HC. According to Figure 4.15.A and B, the FP signals were decreased by reduction of the concentration of PAC-BDP at 1 μM and 25 nM. Curiously, the FP signals at 100 nM did not follow a dose-dependent behavior and this singularity is seen in both experiments. Table 4.1 summarized the biological activity of the ligands that have been tested against BoNTA HC and BoNTA LC in our studies.

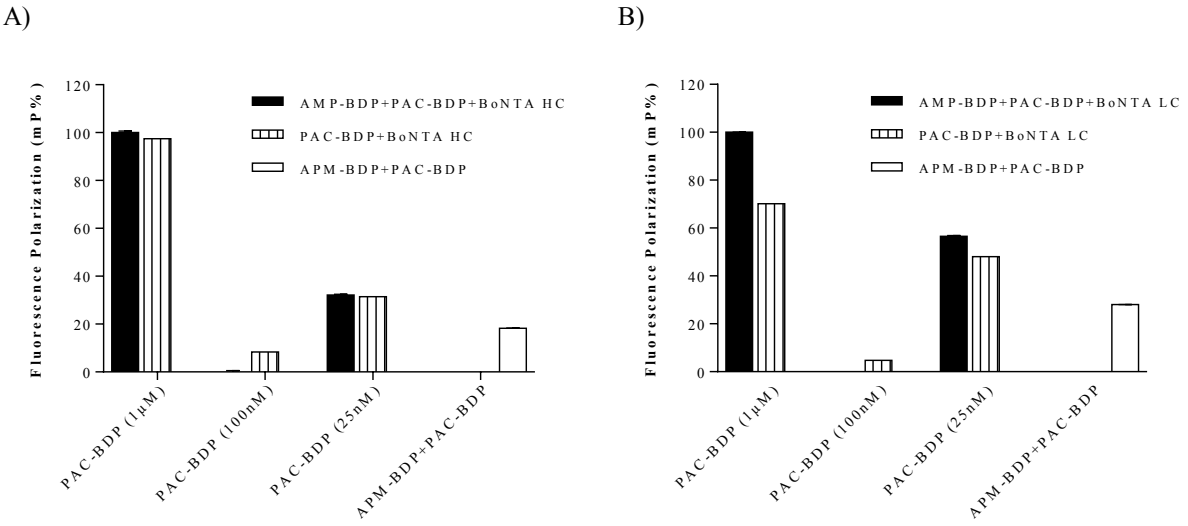


Figure 4.15 Competition assays of APM-BDP and PAC-BDP against (A) BoNTA HC and (B) LC using 570 mirror set. In both experiments the concentrations of APM-BDP (25 nM) and BoNTA (HC or LC) (2 nM) were fixed and the concentration of PAC-BDP was varied. The black bars represent the sample with both fluorescent compounds and BoNTA. The stripe pattern bars correspond to the mixture of PAC-BDP and BoNTA. The white bar represents the baseline FP corresponding to only APM-BDP and PAC-BDP (no toxin). The 570 nm mirror, 540/25 nm excitation filter and 590/35 nm emission filter were used for both FP competition assays.

Table 4.1 Binding profiles of the ligands tested against BoNTA HC and BoNTA LC in this study.

Ligand	Concentration tested	BoNTA LC		BoNTA HC	
		Binding (+)/no binding (-)	EC_{50}	Binding (+)/no binding (-)	EC_{50}
6-AFLU	50 µM	-	-	+	546 ± 60 nM
	1 µM	-		+	
	100 nM	-		+	
	100 pM	-		-	
APM-BDP	1 µM	+	20.96 ± 10 nM	-	-
	100 nM	+		-	
	1 nM	-		-	
	100 pM	-		-	
PAC-BDP	1 µM	+	[1]	+	[1]
	100 nM	+		+	
	25 nM	+		+	
PAC	1 µM	+	16.91 nM[16]		
	20 nM	+			
DES	6 µM	-	-	+	1.6 ± 0.3 µM [3]
	3 µM				
	2 µM			+	
AMN	3 µM	-	-	+	703 ± 98 nM [3]
	0.5 µM			+	
APM	30 µM	+	[2]	_	
	1 µM	+		_	

[1]Enough amount of PAC-BDP was not available at the time when the experiment was carried out. [2]Radiolabeled APM was not available commercially to determine the EC_{50} of APM using scintillation assay. [3] The EC_{50} obtained from other study presented in chapter 5.

There are many examples in the literature of ligand binding studies that employ fluorescent based techniques. In 2011 a FRET (Foster resonance energy transfer) and DARET

(Depolarization After Resonance Energy Transfer) techniques were developed to detect BoNT serotypes in different studies [15, 20]. Recently, FP was used as a high throughput screening assay to identify coumestrol as an inhibitor for the estrogen receptors[21]. In another study the novel rhomboid inhibitors were identified by employing an FP assay[22]. A new antagonist was identified for A(2A) adenosine receptors using FP assay [23]. The FP assay format has some advantages compared to other assay platforms, including the simplicity of sample preparation (mix and read) and the high sensitivity of the assay. Additionally, FP does not require any washing or separating steps which make the assay faster. The FP samples can be miniaturized to lower volumes which save material, and also enable large-scale testing in a high throughput screening assay[24]. The fluorescent conjugate probes APM-BDP and 6-AFLU both exhibit EC_{50}'s in the nanomolar range against BoNTA LC and HC, respectively. The binding affinity of APM-BDP to BoNTA LC was confirmed in the FP specific binding curve. Moreover, the FP competition result of APM-BDP, PAC, APM and PAC-BDP against BoNTA LC supported the specific binding interaction of APM-BDP to BoNTA LC. The binding curve for 6-AFLU, and the FP competition results of 6-AFLU, DES, AMN, PAC-BDP, and APM against BoNTA HC verified the binding of 6-AFLU to BoNTA HC. The FP competition results for PAC-BDP and APM-BDP confirm that PAC-BDP binds to BoNTA HC and BoNTA LC, while APM-BDP showed affinity only to BoNTA LC. These results were confirmed in two individual FP assays using two distinct sets of mirror and filters.

6-AFLU is also known as a fluoresceinamine isomer II. The use of fluoresceinamine isomer II in biochemical assays has not been recently reported. However the isomer I (5-aminofluorescein) was utilized in a variety of research studies for labeling peptides and fatty acids to characterize them by a fluorescence-based assay [25]. Additionally, conjugate 5-

aminofluorescein-albumin has been used as a biomarker in brain tumor surgery [26]. APM is a non-saccharide sweetener produced by coupling L-Aspartic acid and L-phenylalanine. APM was recently found to exhibited binding to DNA of native calf [27].

4.6 CONCLUSIONS

Our results indicate that computational screening is an effective method for identifying recognition agents that bind to spatially distinct site within a target protein. We used a combination of computational and experimental methodologies to identify 6-AFLU and APM-BDP as BoNTA fluorescent recognition agents. 6-AFLU exhibited binding affinity to BoNTA HC whereas APM-BDP showed binding to the catalytic domain located in the BoNTA LC. Binding of fluorescent probes to different locations within the BoNTA structure makes these recognition agent candidates suitable for use in a multi-probes fluorescent polarization assay to detect BoNTA. PAC-BDP, identified in our previous study as a potential recognition agent for BoNTA,[16] and APM-BDP were tested in a competition assay against BoNTA LC, confirming that they bind to different binding site within BoNTA LC. This FP assay also confirmed that using multiple probes to detect BoNTA (complex structure) is feasible.

Many studies have been conducted to identify adequate antitoxin for BoNT. However the antitoxins must be available in the body at the initial phase before the entrance of toxin to the nerve cell. The development of a homogeneous, fast, simple and reliable technique to detect BoNT is necessary, since early diagnostic of botulism is essential for effective treatment [7]. Additionally, such assay should be suitable for high throughput screening to facilitate large-scale implementation in case of bioterrorism [5]. In this study we were able to successfully identify two novel fluorescent recognition agents for BoNTA using PSVLS and FP techniques. A multi-probe FP assay can potentially be developed, considering that the identified fluorescent recognition

agents bind to different binding sites of the BoNTA complex (HC + LC). The development of a multi-probe FP assay for BoNTA complex can become an effective method to diagnose botulism by detecting the presence of the toxin in body samples at early stage. Such assay would also have application in food safety, to detect the presence of the toxin in contaminated food. The active site of BoNTA complex is not accessible for targeting before cleavage of the toxin (activation) into BoNTA LC and BoNTA HC. Therefore the multi-probes FP assay can be an adequate method since these recognition probes targeting the other binding sites (not active site) of BoNTA. The FP assay was found to be rapid, relatively uncomplicated and cheaper to operate compare to the MLA, which is currently the only standard available to detect BoNTA. The detection limit of MLA reported to be 0.01 ng/ml[10]. Based on the result provided in this study, the FP technique was capable of detecting BoNTA LC and BoNTA HC at the concentration of 2 nM, corresponding to 100 and 102 ng/ml, respectively. Although the minimum concentration of BoNT for FP detection obtained in this study is higher than the MLA detection limit, the results are based on individual BoNTA LC or HC. Thus, the sensitivity of the assay needs to be investigated using BoNTA complex. Additionally, lower concentrations (< 100 ng/ml) of BoNTA LC, HC and complex have to be tested for determination of FP threshold.

4.7 ACKNOWLEDGEMENTS

This research was supported with funds from the Thunder Bay Regional Research Institute and the RBC Royal Bank's Dr. Mark Poznansky Mentorship Development Award. High-throughput computational work in this project was made possible with resources provided through SHARCNET (www.sharcnet.ca) and Lakehead University's High Performance Computing Centre (LUHPCC).

4.8 ASSOCIATED CONTENT

Supporting information. The following data are provided as supplementary material: the plots of holistic and force field scores of unlabeled and fluorescent labeled ligands at critical regions of BoNTA, PCA scatter plots of molecular docking of unlabeled and fluorescent labeled libraries for BoNTA, NMR spectrum of APM-BDP, LC/MS spectrum of APM-BDP. This material is available free of charge at http://pubs.acs.org.

4.9 AUTHOR INFORMATION

Corresponding author. Wely Floriano, Lakehead University, Department of Chemistry, 955 Oliver Rd Thunder Bay ON P7B 5E1, Canada, Email: wely.floriano@lakeheadu.ca, Phone: (807) 766-7215.

4.10 ABBREVIATIONS

BoNTA, botulinum neurotoxin serotype A; BoNTA LC, botulinum neurotoxin serotype A light chain; BoNTA HC, botulinum neurotoxin serotype A heavy chain; 6-AFLU, 6-aminofluorescein; APM, aspartame; DES, desmosine; AMN, aminopterin; PAC, paclitaxel; APM-BDP, aspartame bodipy; PAC-BDP, paclitaxel bodipy; FP, fluorescence polarization; DMSO, dimethyl sulfoxide; LC/MS, liquid chromatography-mass spectrometry; NMR, nuclear magnetic resonance; PCA, principal component analysis; MOE, molecular operating environment; PAD, primary amines database; SNAR proteins, soluble NSF-attachment protein receptor.

4.11 REFERENCES

1. Arnon, S. S.; Schechter, R.; Inglesby, T. V.; Henderson, D. A.; Bartlett, J. G.; Ascher, M. S.; Eitzen, E.; Fine, A. D.; Hauer, J.; Layton, M.; Lillibridge, S.; Osterholm, M. T.; O'Toole, T.; Parker, G.; Perl, T. M.; Russell, P. K.; Swerdlow, D. L.; Tonat, K., Botulinum toxin as a biological weapon: medical and public health management. *JAMA : the journal of the American Medical Association* **2001,***285* (8), 1059-70.

2. Cai, S.; Singh, B. R.; Sharma, S., Botulism diagnostics: from clinical symptoms to in vitro assays. *Critical reviews in microbiology* **2007,***33* (2), 109-25.

3. Ouimet, T.; Duquesnoy, S.; Poras, H.; Fournie-Zaluski, M. C.; Roques, B. P., Comparison of fluorigenic peptide substrates PL50, SNAPTide, and BoTest A/E for BoNT/A detection and quantification: exosite binding confers high-assay sensitivity. *Journal of biomolecular screening* **2013,***18* (6), 726-35.

4. Levy, N. S.; Lowenthal, D. T., Application of botulinum toxin to clinical therapy: advances and cautions. *American journal of therapeutics* **2012,***19* (4), 281-6.

5. Dhaked, R. K.; Singh, M. K.; Singh, P.; Gupta, P., Botulinum toxin: bioweapon & magic drug. *The Indian journal of medical research* **2010,***132*, 489-503.

6. Boldt, G. E.; Kennedy, J. P.; Hixon, M. S.; McAllister, L. A.; Barbieri, J. T.; Tzipori, S.; Janda, K. D., Synthesis, characterization and development of a high-throughput methodology for the discovery of botulinum neurotoxin a inhibitors. *Journal of combinatorial chemistry* **2006,***8* (4), 513-21.

7. Cai, S.; Lindo, P.; Park, J. B.; Vasa, K.; Singh, B. R., The identification and biochemical characterization of drug-like compounds that inhibit botulinum neurotoxin serotype A

endopeptidase activity. *Toxicon : official journal of the International Society on Toxinology* **2010,***55* (4), 818-26.

8. Wu, H. C.; Huang, Y. L.; Lai, S. C.; Huang, Y. Y.; Shaio, M. F., Detection of Clostridium botulinum neurotoxin type A using immuno-PCR. *Letters in applied microbiology* **2001,***32* (5), 321-5.

9. Sharma, S. K.; Eblen, B. S.; Bull, R. L.; Burr, D. H.; Whiting, R. C., Evaluation of lateral-flow Clostridium botulinum neurotoxin detection kits for food analysis. *Applied and environmental microbiology* **2005,***71* (7), 3935-41.

10. Lindstrom, M.; Korkeala, H., Laboratory diagnostics of botulism. *Clinical microbiology reviews* **2006,***19* (2), 298-314.

11. (a) Peterson, K. J.; Sadowsky, J. D.; Scheef, E. A.; Pal, S.; Kourentzi, K. D.; Willson, R. C.; Bresnick, E. H.; Sheibani, N.; Gellman, S. H., A fluorescence polarization assay for identifying ligands that bind to vascular endothelial growth factor. *Analytical biochemistry* **2008,***378* (1), 8-14; (b) Leopoldo, M.; Lacivita, E.; Berardi, F.; Perrone, R., Developments in fluorescent probes for receptor research. *Drug discovery today* **2009,***14* (13-14), 706-12.

12. Seethala, R.; Golla, R.; Ma, Z.; Zhang, H.; O'Malley, K.; Lippy, J.; Cheng, L.; Mookhtiar, K.; Farrelly, D.; Zhang, L.; Hariharan, N.; Cheng, P. T., A rapid, homogeneous, fluorescence polarization binding assay for peroxisome proliferator-activated receptors alpha and gamma using a fluorescein-tagged dual PPARalpha/gamma activator. *Analytical biochemistry* **2007,***363* (2), 263-74.

13. Parker, G. J.; Law, T. L.; Lenoch, F. J.; Bolger, R. E., Development of high throughput screening assays using fluorescence polarization: nuclear receptor-ligand-binding and kinase/phosphatase assays. *Journal of biomolecular screening* **2000,***5* (2), 77-88.

14. (a) Jameson, D. M.; Croney, J. C., Fluorescence polarization: past, present and future. *Combinatorial chemistry & high throughput screening* **2003,***6* (3), 167-73; (b) Lee, P. H.; Bevis, D. J., Development of a homogeneous high throughput fluorescence polarization assay for G protein-coupled receptor binding. *Journal of biomolecular screening* **2000,***5* (6), 415-19.

15. (a) Gilmore, M. A.; Williams, D.; Okawa, Y.; Holguin, B.; James, N. G.; Ross, J. A.; Roger Aoki, K.; Jameson, D. M.; Steward, L. E., Depolarization after resonance energy transfer (DARET): a sensitive fluorescence-based assay for botulinum neurotoxin protease activity. *Analytical biochemistry* **2011,***413* (1), 36-42; (b) Ross, J. A.; Gilmore, M. A.; Williams, D.; Aoki, K. R.; Steward, L. E.; Jameson, D. M., Characterization of Forster resonance energy transfer in a botulinum neurotoxin protease assay. *Analytical biochemistry* **2011,***413* (1), 43-9.

16. Dadgar, S.; Ramjan, Z.; Floriano, W. B., Paclitaxel Is an Inhibitor and Its Boron Dipyrrornethene Derivative Is a Fluorescent Recognition Agent for Botulinum Neurotoxin Subtype A. *J Med Chem* **2013,***56* (7), 2791-2803.

17. Larkin, M. A.; Blackshields, G.; Brown, N. P.; Chenna, R.; McGettigan, P. A.; McWilliam, H.; Valentin, F.; Wallace, I. M.; Wilm, A.; Lopez, R.; Thompson, J. D.; Gibson, T. J.; Higgins, D. G., Clustal W and Clustal X version 2.0. *Bioinformatics* **2007,***23* (21), 2947-8.

18. Floriano, W. B.; Vaidehi, N.; Zamanakos, G.; Goddard, W. A., 3rd, HierVLS hierarchical docking protocol for virtual ligand screening of large-molecule databases. *J Med Chem* **2004,***47* (1), 56-71.

19. Hammer, Ø.; Harper, D. A. T., *Paleontological data analysis*. Blackwell Pub.: Malden, MA, 2006; p x, 351 p.

20. Ruge, D. R.; Dunning, F. M.; Piazza, T. M.; Molles, B. E.; Adler, M.; Zeytin, F. N.; Tucker, W. C., Detection of six serotypes of botulinum neurotoxin using fluorogenic reporters. *Analytical biochemistry* **2011,***411* (2), 200-9.

21. Wang, C.; Li, C.; Zhou, H.; Huang, J., High-Throughput Screening Assays for Estrogen Receptor by Using Coumestrol, a Natural Fluorescence Compound. *Journal of biomolecular screening* **2013**.

22. Wolf, E. V.; Zeissler, A.; Vosyka, O.; Zeiler, E.; Sieber, S.; Verhelst, S. H., A new class of rhomboid protease inhibitors discovered by activity-based fluorescence polarization. *PloS one* **2013,***8* (8), e72307.

23. Kecskes, M.; Kumar, T. S.; Yoo, L.; Gao, Z. G.; Jacobson, K. A., Novel Alexa Fluor-488 labeled antagonist of the A(2A) adenosine receptor: Application to a fluorescence polarization-based receptor binding assay. *Biochemical pharmacology* **2010,***80* (4), 506-11.

24. (a) Pope, A. J.; Haupts, U. M.; Moore, K. J., Homogeneous fluorescence readouts for miniaturized high-throughput screening: theory and practice. *Drug discovery today* **1999,***4* (8), 350-362; (b) Lea, W. A.; Simeonov, A., Fluorescence polarization assays in small molecule screening. *Expert opinion on drug discovery* **2011,***6* (1), 17-32.

25. (a) Uryga-Polowy, V.; Kosslick, D.; Freund, C.; Rademann, J., Resin-bound aminofluorescein for C-terminal labeling of peptides: high-affinity polarization probes binding to polyproline-specific GYF domains. *Chembiochem : a European journal of chemical biology* **2008,***9* (15), 2452-62; (b) Brando, T.; Pardin, C.; Prandi, J.; Puzo, G., Analysis of aminofluorescein-fatty acid derivatives by capillary electrophoresis with laser-induced fluorescence detection at the attomole level: application to mycobacterial fatty acids. *Journal of chromatography. A* **2002,***973* (1-2), 203-10.

26. Ding, R.; Frei, E.; Fardanesh, M.; Schrenk, H. H.; Kremer, P.; Haefeli, W. E., Pharmacokinetics of 5-aminofluorescein-albumin, a novel fluorescence marker of brain tumors during surgery. *Journal of clinical pharmacology* **2011,***51* (5), 672-8.

27. Kashanian, S.; Khodaei, M. M.; Kheirdoosh, F., In vitro DNA binding studies of Aspartame, an artificial sweetener. *Journal of photochemistry and photobiology. B, Biology* **2013,***120*, 104-10.

5 CHAPTER FIVE: TARGETING NON-ORTHOSTERIC SITES IN PROTEINS IS A VALID STRATEGY FOR TARGET-SPECIFIC MOLECULAR PROBE DISCOVERY

Saedeh Dadgar and Wely B. Florino[*]

Will be submitted to the *J. Med. Chem.* in January 2014

Department of Chemistry at Lakehead University and Thunder Bay Regional Research Institute,

Thunder Bay ON P7B 5E1, Canada

AUTHOR EMAIL ADDRESS: wely.floriano@lakeheadu.ca

TITLE RUNNING HEAD: Target-specific, non-orthosteric, molecular probes

5.1 PREFACE

In this study, we further analyzed the computational results reported in chapter 3 in order to identify non-orthosteric molecular probes for BoNTA. The specific binding of Desmosine (1.6 ± 0.3μM) and Aminopterin (703 ± 98 nM) to BoNTA LC was determined using scintillation proximity assay. Solanesyl pyrophosphate exhibited binding affinity to both BoNTA HC and LC. These recognition probes were selected to target the non-conserved regions in BoNTA and may show selectivity for BoNTA over the other subtypes.

5.2 ABSTRACT

We previously reported the use of Protein Scanning with Virtual Ligand Screening (PSVLS) to identify a novel inhibitor and a recognition agent (molecular probe) against the catalytic site of the *botulinum* neurotoxin subtype A. In the present work, we further analyze the PSVLS data to identify molecular probes with affinity for binding regions other than the catalytic site of BoNTA. We refer to these as non-orthosteric molecular probes. Such probes can take advantage of unique characteristics within the structure of target proteins that discriminate them from other closely related proteins. Molecular probes targeting binding sites that are unique to the serotype of interest are expected to be less likely to bind to other serotypes, making a detection assay developed with such probes less prone to false positives. BoNTA is a large protein with a light (LC) and a heavy (HC) chain that can be assayed separately. The catalytic site is located in the LC. We selected for experimental testing 3 probe candidates predicted computationally to have affinity for different non-orthosteric binding regions within the HC and LC. One compound predicted not to have affinity for either domain was also selected for testing. Two previously tested compounds were used as positive and negative controls. The selected compounds were purchased with a tritium label and tested experimentally for binding to either

the LC or the HC of BoNTA using scintillation proximity assay (SPA). The binding profiles obtained experimentally confirmed that our computational methodology can successfully identify probe candidates targeting multiple and spatially distinct binding pocket within the target protein. Non-orthostheric ligands provide new opportunities for controlling the function of BoNTA without direct catalytic inhibition. They may also serve as subtype-specific molecular probes that take opportunity of unique binding pockets in BoNTA compared to other subtypes or other homologous proteins (potential confounders in a detection assay). Moreover, our previously reported inhibition assay results for these probe candidates indicate that *PSVLS* can identify probe candidates targeting binding sites that do not significantly interfere with the catalytic activity of the target protein. The identification of non-interfering probes opens up new opportunities for molecular imaging probe discovery and development. Non-interfering molecular imaging probes can be used for monitoring novel disease therapies without the risk of skewing the results with effects attributed to the probe itself.

KEYWORDS: molecular probes, radiolabeled probes, non-orthosteric binding, binding to alternative sites in proteins, detection agents, *botulinum* neurotoxin subtype A, virtual screening, holistic binding scoring function, drug discovery.

BRIEFS: Novel detection agents targeting non-orthosteric sites in botulinum neurotoxin subtype A were discovered by virtual screening and confirmed experimentally.

5.3 INTRODUCTION

A primary (orthosteric) ligand binds to the orthosteric site of its cognate protein to induce or suppress response. Traditionally drug discovery have targeted orthosteric sites in proteins as a way to control their activity. However, this approach fails to produce ligands with enough subtype specificity because of the high degree of conservation generally associated to functionally active sites. Since subtype specificity is important for many pharmacologically relevant targets such as G protein coupled receptors, alternative approaches have been sought. Recently, structure-based drug discovery efforts have shifted towards the targeting of allosteric sites as a way to bypass the high degree of conservation of functional sites among protein subtypes and closely related homologs[1]. Allosteric sites are spatially distinct from the orthosteric site. Binding of a ligand to an allosteric site induces conformational changes that affect the binding of the orthosteric ligand and/or the response of the target protein. Allosteric sites are still functional sites in the sense that they modulate the primary function of the protein. However they are under less evolutionary pressure to be conserved than orthosteric sites and hence their appeal as target sites for subtype specificity. Conceptually, other binding pockets within a protein may exist that are large enough to accommodate a ligand without producing a significant change in the activity associated to an orthosteric site. Such sites may provide unique characteristics that allow the discrimination of a protein from other closely related proteins. This type of discrimination is especially important in the context of target-specific molecular probes (TS-MPs), which are recognition agents that detect a biomolecule of interest. TS-MPs are usually obtained by labeling a known ligand with a radioisotope or a fluorophore. Because they are based on known ligands, TS-MPs usually interfere with the function of the protein they are targeting, which prevents their use to monitor the biological effects of novel ligands. Moreover,

they are often plagued by the same specificity problems their originating ligand may have. In the context of TS-MPs, the possibility of finding specific probes for non-interfering/non-orthosteric sites is very attractive and may represent a shift in the current paradigm of probe discovery. A number of key questions, however, remain to be answered: would ligands targeting these sites have enough specific affinity for the target protein to enable their use as molecular probes? Wouldn't the binding of any ligand, anywhere within a protein alter its response to an orthosteric ligand? In order to gain insights into these questions, we targeted non-orthosteric sites in the *botulinum* neurotoxin subtype A structure for the identification of molecular probe candidates.

Botulinum neurotoxin (BoNT), one of the most known lethal toxins and a causative agent of botulism disease, is produced by the bacterium *Clostridium botulinum*[2]. Botulism is a serious disease which affects the peripheral nerve system by cleavage of a group of proteins (SNARE). These proteins are involved in membrane fusion and their cleavage consequently inhibits acetylcholine secretion resulting in respiratory muscles paralysis and ultimately death [3]. There are seven types of BoNT classified by the letters A - G and BoNTA is known as the most potent one [4]. BoNT is a polypeptide molecule with a molecular weight of 150 KD. Activated BoNT is cleaved in two smaller chains, a heavy (HC) and a light chain (LC) with molecular weights of 100 and 50 KD, respectively. The active site of BoNT with zinc-endopeptidase activity is located in the LC (catalytic domain) whereas the HC contains the binding domain of BoNT [4-5]. BoNTA and B have been used as a drug since 2008 [6]. In recent years, several studies have attempted to identify new inhibitors of BoNT due to its various applications in different areas such as medical, pharmaceutical, cosmetic and the potential threat of using BoNT as a biological weapon [6-7]. However, most of these studies focused on identification of novel recognition agents and inhibitors for the active site of BoNTA. Our group has also targeted the active site of BoNTA [8],

successfully identifying a novel inhibitor and a fluorescent recognition agent. In that work, a library consisting of 1624 compounds including commercially available radiolabeled ligands was computationally screened against BoNTA. The PSVLS (Protein Scanning with Virtual Ligands Screening) method was used for the virtual screening. In the PSVLS method, the binding affinity of each ligand for every individual binding site available in BoNTA is calculated and then ligands are ranked based on their holistic scores [8]. The holistic scores take into account the binding affinity of the ligand for all the other regions relative to the binding region of interest (in that paper, the catalytic site). The outcome of the PSVLS/holistic binding scoring was evaluated to determine novel inhibitors for BoNTA's catalytic site.

Here we report the analysis of the PSVLS results in Dadgar et al., 2013 [8] for all 33 binding regions explored by PSVLS, with the intent of identifying non-orthosteric ligands for BoNTA. Non-orthostheric ligands may serve as subtype-specific molecular probes that take opportunity of unique binding pockets in BoNTA compared to other subtypes or other homologous proteins (potential confounders in a detection assay). They also provide new opportunities for controlling the function of BoNTA without direct catalytic inhibition. We selected four ligands based on their PSVLS binding profiles. Our previously identified inhibitor, paclitaxel, was used as positive control. D-fructose, predicted to be a non-binder by PSVLS which was confirmed in our previous work, was used as negative control. Selected ligands were tested against BoNTA HC and BoNTA LC separately using scintillation proximity assay (SPA)[8]. [^{3}H] aminopterin and [^{3}H] desmosine exhibited binding interactions to BoNTA HC however considering the error bars no significant interactions observed against BoNTA LC. [^{3}H] solanesol showed no binding affinity for both BoNTA LC and HC however [^{3}H] solanesyl pyrophosphate exhibited high binding interaction against both BoNTA LC and HC. Our results demonstrate that PSVLS/holistic

binding scoring approach is an accurate and reliable method to identify molecular probe candidates targeting orthosteric and non-orthosteric sites in BoNTA.

5.4 METHODS AND PROCEDURES

5.4.1 Protein Scanning with Virtual Ligand Screening (PSVLS).

This method was used in our previous work [8] to screen a virtual library of 1,624 ligands against BoNTA in order to identify new inhibitors and molecular probes targeting the catalytic site of BoNTA, a metalloprotease. The results of that screening for all 33 binding sites of BoNTA are being further analyzed in the present work with the intent of identifying non-orthosteric ligands for BoNTA. PSVLS is a computational protocol for virtual screening that targets multiple sites within the 3-dimensional (3D) structure of the target protein. PSVLS takes the 3D structures of a target protein and of a library of chemical compounds, scans the protein for potential binding sites and screens the virtual ligand library for chemical compounds with high affinity for any of the potential binding sites [9]. The core computational methodology used in PSVLS is HierVLS[9a], a multi-step hierarchical protocol for virtual ligand screening. HierVLS is a computational protocol for fast screening of large (> 500,000) chemical compounds library [9a]. HierVLS uses a hierarchical approach: a large number of bound protein-ligand configurations are created in the least computationally expensive step for each ligand. Subsequent steps reduce the number of bound configurations per ligand while increasing the computational time required to process them (for increased accuracy). The last step is the calculation of the binding energy including solvation for the best surviving bound configuration of each ligand. Binding energies are calculated using a force field-based scoring function. For PSVLS, we apply HierVLS not only to the active site of the target protein, but also to other available pockets, which can be used

for detection of the protein or as alternative sites for controlling biological activity. These alternative binding sites are identified as enclosed empty volumes large enough to accommodate small molecules. PSVLS provides binding affinities and bound structures for each ligand in the screening library, in each one of the potential binding sites in the target protein. The centers of the 33 binding sites identified in BoNTA (pdb code 2NYY) are shown in Figure 5.1 as solid spheres, with the HC and LC domains of BoNTA labeled.

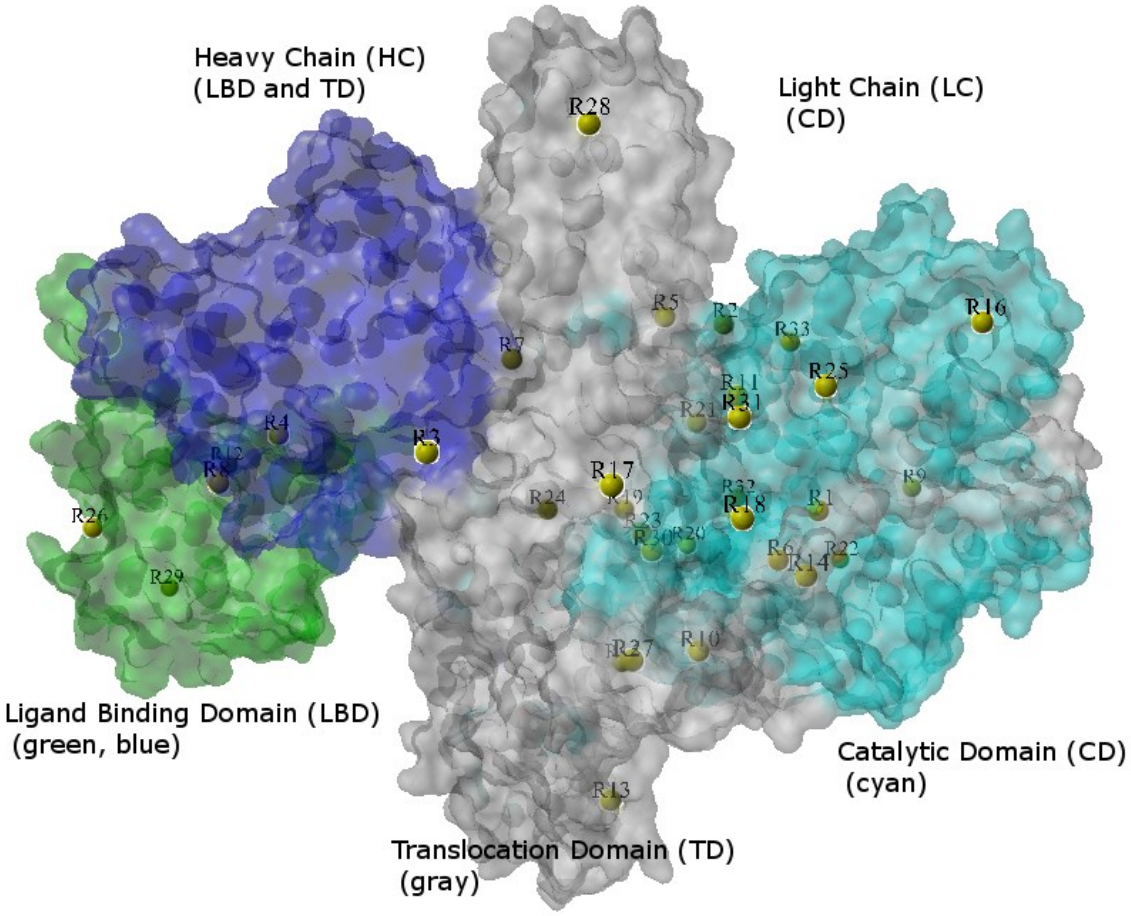


Figure 5.1 Molecular surface of BoNTA (pdb code 2NYY) with the centers of each of 33 binding regions represented as spheres. The light chain (LC) and heavy chain (HC) domains which can be assayed independently of each other are highlighted in the figure.

5.4.2 Selection of Probe candidates for experimental testing

Principal component analysis (PCA) was used to identify ligands likely to bind experimentally to BoNTA. This approach takes into account the affinity of the probe candidate for the target protein as a whole. PCA is used to reduce the binding energies across all regions into a single predictive score for each chemical compound screened. The first principal component (PC1), which has the largest variance, represents the ligands' overall affinity for the protein. A non-binder/binder threshold is set as one standard deviation below the mean value (zero for the transformed data) of the entire screening library. The underlying assumption is that the screening of a random library will mostly produce scores that represent non-specific interactions between the ligands and the protein, with only a small number of compounds displaying "true" binding interactions and thus passing the threshold. Ligands with a PC1 score passing the non-binder/binder threshold are further analyzed in terms of their individual holistic scores[8] at each binding region. This allows for the selection of ligands that preferentially bind to specific sites within the protein. Practical aspects such as commercial availability with a tritium label, price, safety, and easy handling are also considered in the selection of ligands for experimental testing among the ones passing the binder/non-binder criteria.

5.4.3 Materials for experiments

BoNTA LC and HC were purchased from List Biological laboratories,inc. Radioligands were obtained from American Radiolabeled Chemicals Inc., except [^{3}H] aminopterin which was purchased from MoravekBiochemicals. The specific activity of [^{3}H] aminopterin(ETOH:H2O 4:6), [^{3}H] desmosine (H2O), [^{3}H] paclitaxel(Ethylacetate), [^{3}H] solanesol(Hexane), [^{3}H] solanesyl pyrophosphate (IPA:NH4OH:H2O)and [^{3}H] fructose-D(Ethanol) were 38.4 Ci/mmol, 5Ci/mmol, 60 Ci/mmol, 20 Ci/mmol, 20 Ci/mmol and 5 Ci/mmol, respectively. The assay buffer

contained 20 mM HEPES buffer at pH 7.4, 0.1% Tween 20, 0.3 mM $ZnCl_2$, andwas used to prepare the dilutions of radioligands. The ChromaLink™Biotin One-Shot Antibody Labeling kit was purchased from Solulink and used to biotinylated BoNTA LC and BoNTA HC. The hydrolysis buffer contained 20 mM HEPES buffer at pH 7.4 and 1%. Tween 20, and was used to make dilutions of BoNTA HC and BoNTA LC. Streptavidin coated PolyVinyltoluene (PVT) scintillation proximity assay (SPA) beads were supplied from PerkinElmer, Inc.

5.4.3.1 Biotinylation of BoNTA HC and BoNTA LC

BoNTA LC (30 µg) and BoNTA HC (50 µg) were biotinylated using the ChromaLink™ Biotin One-Shot Antibody Labeling Kit according to Solulink's instruction. To determine the concentration of the proteins, we needed two parameters including the mass extinction coefficient (E1%) for BoNTA/HC and LC and also their absorbance at 280 nm. The E1% of BoNTA/HC was reported to be 17.1 by Weatherly, G. T. et al [10]. E1% value of BoNTA/LC was not available in literature. According to the extinction coefficient technical guide which exists on the Thermo scientific website, if the E1% of any protein is not available, then E1% equals 10 can be used as a rough estimation value for the protein. The biotin molar substitution ratio of biotinylated BoNTA LC and biotinylated BoNTA HC was calculated to be 2 and 4.3, respectively, using the MSR calculator of Solulink. The total biotin-labeled BoNTA LC recovered after biotinylation was 28.3 µg and the total labeled BoNTA HC was equal to 43.9 µg.

5.4.3.2 Scintillation proximity assay

The order of adding reagents was optimized in the assay as described below. Radioligand and protein (BoNTA LC or HC) were incubated for 30 minutes at room temperature to pre-equilibrate before the addition of SPA beads to decrease the chance of non-specific binding. The SPA beads were reconstituted in assay buffer with a concentration of 50 mg/ml and stored at 4

ºC. SPA beads working solutions at 5 mg/ml were prepared on the day of each experiment. The concentration of SPA beads and biotinylated BoNTA HC (or BoNTA LC) were fixed in all samples at 2 mg/ml and 17 nM (or 10 nM), respectively. All the experiments were done in 96 well, white and flat bottom Wallac microplates and each sample was run in duplicate at room temperature. Scintillation was counted using a MicroBeta^2 plate counter. To obtain comparable data from different runs, a correction for the radioactive half-life was applied automatically by the MicroBeta^2 plate counter.

The scintillation proximity assay was performed to measure the binding of five different radioligands to BoNTA HC (or LC). [^{3}H] paclitaxel was used as a positive control at 50 nM concentration and [^{3}H] fructose was used as a negative control at the concentration of 3 μM. These concentrations were chosen based on our previous experiments [8]. Non-specific binding of radioligands to the beads was determined in the absence of BoNTA. Specific binding was calculated by subtracting non-specific binding from total binding. Similarly, the binding interaction of radioligands was measured for BoNTA HC. [^{3}H] aminopterin, [^{3}H] desmosine, [^{3}H] solanesol, and [^{3}H] solanesyl pyrophosphate were initially tested at the concentrations of 1 μM, 2 μM, 300 nM, and 600 nM, against BoNTA HC (or LC). Additionally, [^{3}H] aminopterin, and [^{3}H] desmosine were assayed at two lower concentrations (250 nM and 50 nM) for binding to BoNTA HC in a separate assay.

Assay format optimization. To develop the SPA for BoNTA, all three formats of adding reagents were tested including: coupling SPA beads and BoNTA before adding Radioligands, adding all three reagents at the same time, and incubating BoNTA with radioligand before the addition of SPA beads. Based on our study, the most suitable format of the SPA for BoNTA is delayed addition of SPA beads which also reduces the non-specific binding.

Chromophore absorbance correction. According to the emission spectrum of SPA beads (PVT) provided in the Amersham SPA product guide, the SPA beads (PVT) emit between 360-490 nm. ChromaLink^{TM} (used in biotinylation of BoNTA LC and HC) has a chromophore with maximum absorbance at 354 nm, which overlaps the emission spectrum of SPA beads (PVT). This means that part of the SPA beads' emitted light may be absorbed by the Chromalink bound to BoNTA. This overlapping explains the high number of controls compared to the samples which was observed in our initial experiments. In this study, the only factor that was different between the control and the sample was the availability of BoNTA in the sample and each sample was compared to its control. To eliminate the effects of the Chromalink, the same amount of Chromalink should be available in each particular sample and its control. According to the data determined after biotinylation of BoNTA HC and LC, the ratio of biotin-Chromalink to BoNTA in labeled BoNTA LC and HC are known so the same concentration of Chromalink was added for the controls.

Spa beads optimization. The ratio of protein and SPA beads is one of the critical factors in the SPA to obtain a maximum signal-noise ratio. In this experiment, the concentration of SPA beads was variable between 0.5 - 2.5 mg/ml while the concentrations of BoNTA HC and [^{3}H] Aminopterin were fixed at 17 nM and 1 µM, respectively. The final volume of each sample was 200 µl and the samples were counted at the speed of 1 well/min. The final concentration of 2 mg/ml of SPA beads was used in our Study since high separation between total binding and non-specific binding was observed at this concentration.

5.4.3.3 [^{3}H] Paclitaxel binding curve

The effective concentration (EC_{50}) of [^{3}H] paclitaxel against BoNTA LC was determined using SPA assay in our previous study. The concentration of BoNTA LC and SPA beads were fixed at 7 nM and 2 mg/ml, respectively [8]. The [^{3}H] paclitaxel concentration was varied from 160- 1nMin this assay.

5.4.3.4 [^{3}H] Aminopterin binding curve

SPA was used to determine the effective concentration (EC_{50}) of [^{3}H] aminopterin against BoNTA HC at 17 nM, with SPA beads at 2 mg/ml, at room temperature. The concentration of [^{3}H] aminopterin varied from 50 nM to 1 μM. Assay buffer was used to dilute [^{3}H] aminopterin for this assay. The assay ran in duplicate at a final volume of 200 μl and the microplate was counted for 8 hours. The data was analyzed using Prism 6 (GraphPad, Inc).

5.4.3.5 [^{3}H] Desmosine binding curve

SPA was used to determine the EC_{50} of [^{3}H] desmosine against BONTA HC. The concentrations of BoNTA HC and SPA beads were kept constant at 17 nM and 2 mg/ml, respectively, while the concentration of [^{3}H] desmosine varied from 10 nM to 4 μM. The [^{3}H] desmosine dilutions were prepared using assay buffer. The samples were counted for 18 hours, but the specific binding remained constant after 10 hours.

5.5 RESULTS AND DISCUSSION

5.5.1 Selection of probe candidates for experimental testing

PCA was applied to the PSVLS results in Dadgar and co-authors (2013) to reduce the force field scores of each ligand bound to each of the 33 binding regions identified in the structure of BoNTA (Figure 5.1) into one score (PC1). Ligands with a PC1 score passing the non-

binder/binder threshold (Figure 5.2) are expected to bind experimentally to BoNTA, and were further analyzed in terms of their individual holistic scores at each binding region. The objective of the PCA was to identify ligands with strong overall affinity for BoNTA to increase the chances of finding true positive compounds. The objective of the per-region analysis was to identify ligands preferentially binding to spatially distinct regions of BoNTA belonging to each of the two domains, HC and LC. These domains are structurally stable and functional as separate units, and can be assayed independently of each other. The PSVLS-predicted binding profiles can thus be tested experimentally.

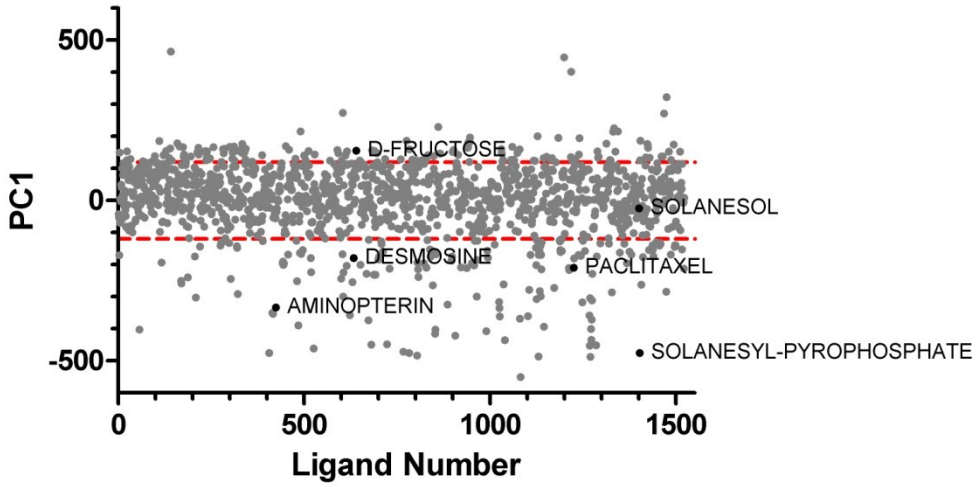


Figure 5.2 PC1 scores for a library of 1,624 compounds computationally screened against BoNTA whole structure (HC and LC). Dashed lines mark one standard deviation above/below the mean score for the library. Ligands selected for experimental testing are highlighted.

In addition to the previously tested [8] paclitaxel (positive control) and fructose (negative control), we selected aminopterin, desmosine, and solanesyl pyrophosphate. All of these compounds passed the non-binder/binder threshold. The region-specific binding energies for these compounds (Figure 5.3) indicated preferential binding to the HC (aminopterin and

desmosine), LC (paclitaxel), or both chains (solanesyl pyrophosphate). Desmosine is a rare amino acid produced by elastin breakdown. The release of desmosine is accelerated in patients with an inflammatory conditions such as chronic obstructive pulmonary disease (COPD)[11], and pulmonary tuberculosis (TB)[12]. Elastin is one of the main extracellular matrix proteins which provide elasticity and flexibility of body tissues. In patients with pulmonary inflammation diseases, the elastin of lung tissue degenerated by interaction of modulators and enzymes released by lymphocytes, neutrophils and microphages [13]. The degradation of elastin produces desmosine and isodesmosine which can be measured in body fluids as a biomarker for the progress of pulmonary disease[14]. Solanesyl pyrophosphate is predicted to be the strongest binder among the ligands selected. As expected, paclitaxel passed the PC1 non-binder/binder threshold, and its region-specific binding energies (Figure 5.3) indicate preferential binding to BoNTA LC, where the catalytic site is located. This is in agreement with previously reported inhibition assays which confirmed paclitaxel as an inhibitor of BoNTA[8]. Also as expected, fructose did not pass the non-binder/binder threshold. Solanesol, a derivative of solanesyl pyrophosphate, was selected for testing to contrast against its parent compound, even though solanesol's PC1 score is higher than the threshold (albeit lower than the average score for the library). Solanesol is extracted from tobacco leaves and has been used as a primary material to synthesize vitamin K and coenzyme Q10 [30]. Aminopterin is a known anticancer drug and has been used to treat leukemia[15]. The ligands selected for experimental testing are listed in Table 5.1 along with their predicted binding profiles.

Table 5.1 Name and structure of the selected ligands from PSVLS. Aminopterin, desmosine, paclitaxel and fructose were tested at 10 µM using FRET (Fluorescent Resonance Energy Transfer) for their inhibition against BoNTA LC (10nM) in our previous paper[8].

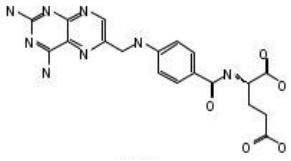
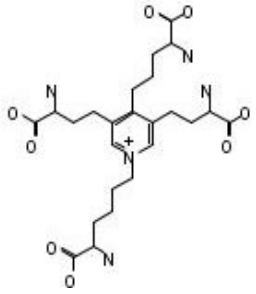
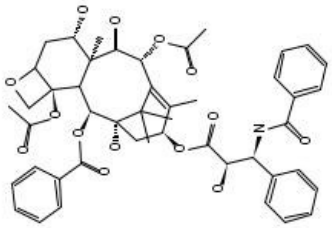
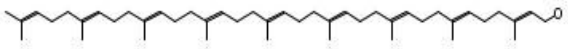
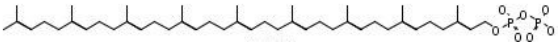
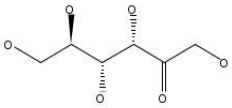
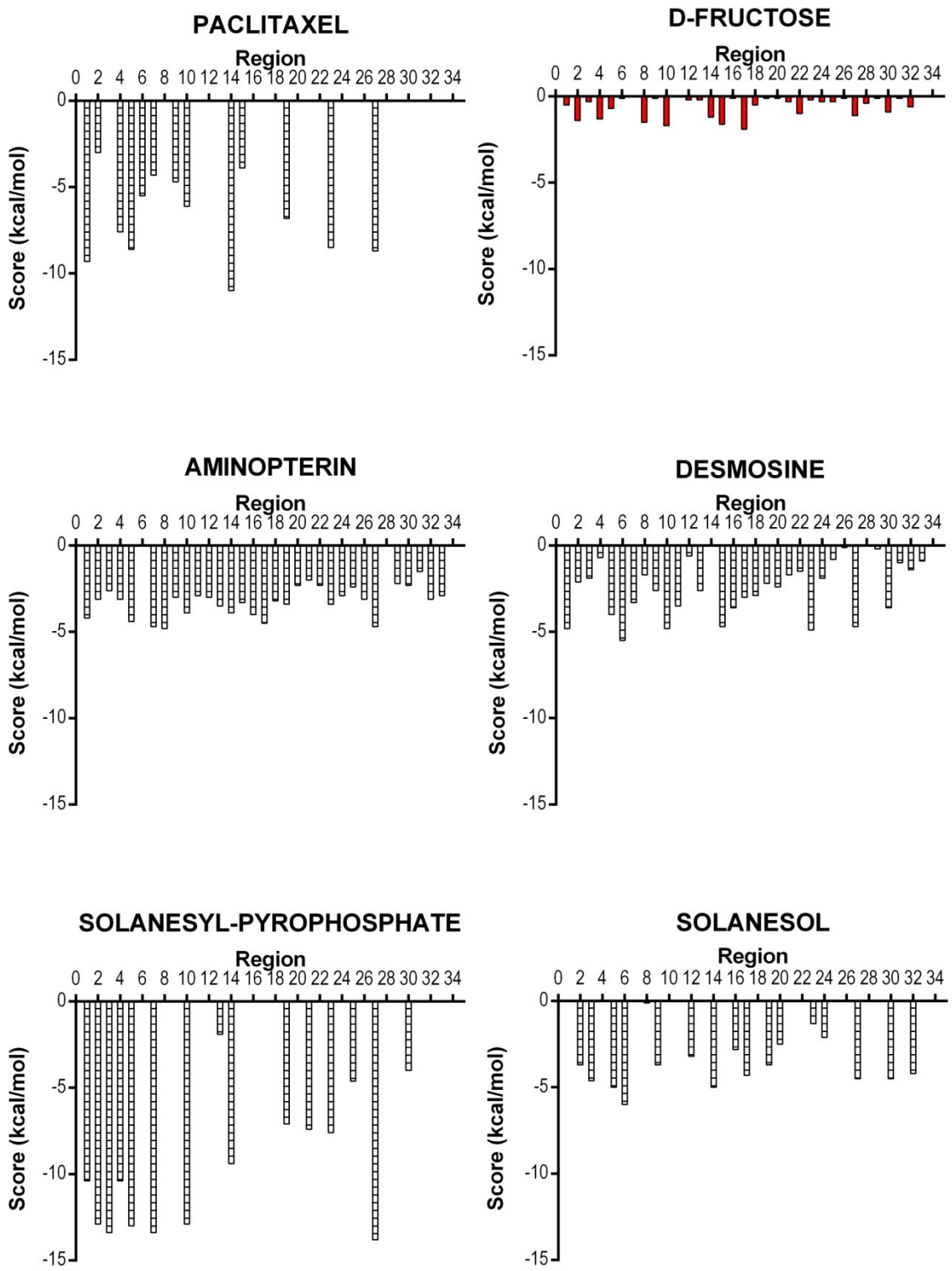
Ligand	% Inhibition at 10 µM	Chemical structure
Aminopterin	11	
Desmosine	25	
Paclitaxel	95	
Solanesol	NT	
Solanesyl pyrophosphate	NT	
D-(-)Fructose	0	

Figure 5.3 Holistic Scores (HS) per region for selected ligands passing the PC1 non-binder/binder threshold.



5.5.2 Optimization of BoNTA and SPA beads ratio

Basically, in a scintillation proximity assay it is necessary to optimize the concentration of SPA beads for each particular protein. The optimum concentration of SPA beads in this particular experiment was identified as 2 mg/ml. The separation between total binding and non-specific binding was relatively high at this concentration of SPA beads (Figure 5.4). In this assay, [^{3}H] aminopterin and BoNTA HC were used for optimizing the concentration of SPA. However, the optimal concentration of SPA beads was also determined to be 2 mg/ml using a separate assay. The final concentrations of [^{3}H] paclitaxel and BoNTA LC were constant in each sample at (6 nM) (10 nM), respectively.

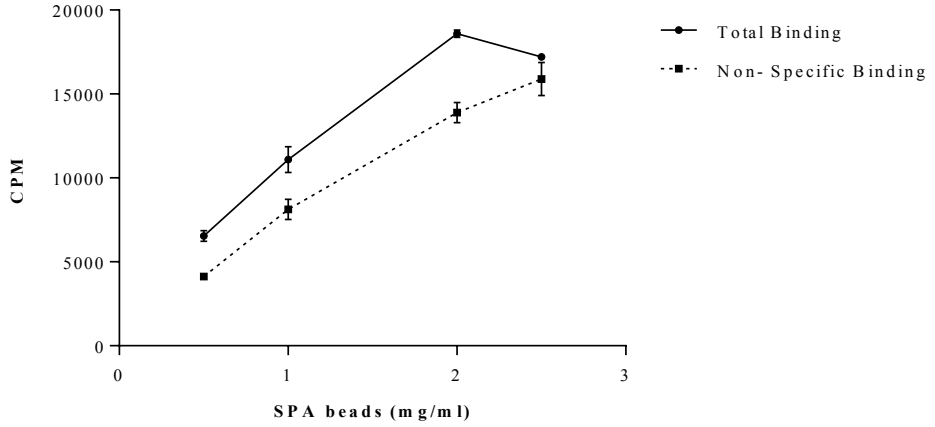


Figure 5.4 Optimization of SPA beads concentration. Total binding (solid line) at 2 mg/ml shows a wider separation from the non-specific binding (dotted line).The total binding and non-specific binding were measured using samples that are exactly the same composition, except for the presence of BoNTA, which is absent in non-specific binding samples.

5.5.3 Experimental binding to BoNTA HC and LC

Preliminary SPA was performed to investigate the binding the selected ligands against BoNTA LC and BoNTA HC. Aminopterin was tested at (1000, 250 and 50 nM) and desmosine

was tested at (2000, 250 and 50 nM) concentrations against BoNTA HC separately. Due to observe no interaction of aminopterin (1000 nM) and desmosine (2000 nM) to BoNTA LC, the lower concentrations were not tested. Solanesol was tested at (1000, 300, 50 and 10 nM) and Sol.P.P was tested at (600, 300, 50 and 10 nM) for BoNTA HC. Sol and Sol.P.P also tested against BoNTA LC only at 600 and 300 nM, respectively. All the data provided in supporting information (Appendix E). A summary of these results is provided in Figure 5.5. The lowest positive concentration tested is provided for positive ligands, whereas the highest concentration tested is shown for negative ligands. [^{3}H] aminopterin and [^{3}H] desmosine both show high affinity for BoNTA HC, but very low binding to BoNTA LC. Accordingly, [^{3}H] solanesyl pyrophosphate exhibits one of the strongest binding interactions not only to BoNTA HC but also to BoNTA LC. The data related to [^{3}H] solanesyl pyrophosphate is shown on the right axis in Figure 5.5. In contrast, [^{3}H] solanesol did not bind to the BoNTA HC or LC at the concentrations tested. [^{3}H] paclitaxel, which was known to be a BoNTA inhibitor in our previous work [8], only shown binding to BoNTA LC at the concentrations tested. In addition, [^{3}H] D-fructose, which was selected as a negative control, does not bind to either BoNTA HC or BoNTA LC.

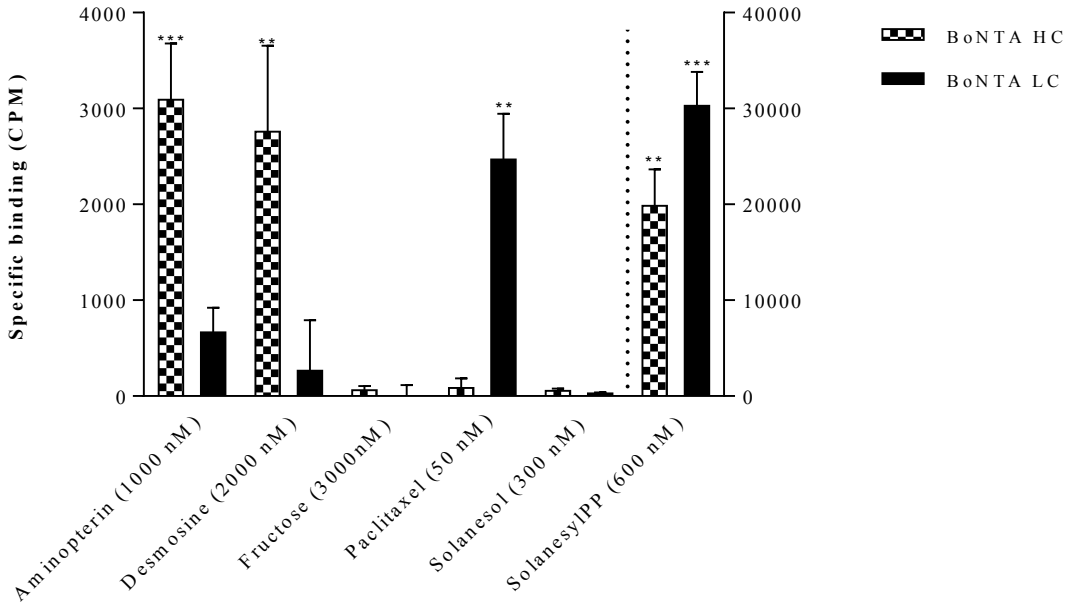


Figure 5.5 Specific binding of each radioligand to BoNTA HC and BoNTA LC using SPA. Binding of radioligands to BoNTA LC or HC are shown by solid black or square pattern, respectively. The dotted line separates Solanesyl pyrophosphate (Solanesyl.PP) from the other radio-ligands due to the fact that Solanesyl.PP is plotted on the right y-axis. The error bars in the graph represent standard deviations and the stars indicate the significant differences between the samples and negative control (fructose) according to two way Anova analysis.

5.5.4 [^{3}H] Paclitaxel binding curve

The [^{3}H] Paclitaxel binding curve against BoNTA LC was determined in our previous study showed in Figure 5.6 [8]. The EC_{50} of paclitaxel was estimated at 16.91 nM for BoNTA LC. The result provided in the graph is after deduction of non-specific binding and normalizing the data[8].

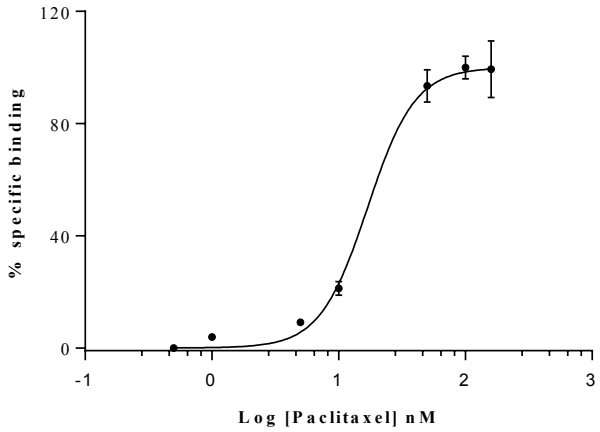


Figure 5.6 SPA binding curve for paclitaxel against BoNTA LC. The EC_{50} of [^{3}H] paclitaxel was determined to be 16.91 nM. The concentration of [^{3}H] paclitaxel varied from 160 to 1 nM, and the concentration of SPA beads and BoNTA LC was constant for all the duplicate samples and controls at 2 mg/ml and 7 nM, respectively. Non-linear regression was used to plot the curve and estimate EC_{50}. The error bars correspond to standard deviations.

5.5.5 [^{3}H] Aminopterine binding curve

The result of the [^{3}H] aminopterin binding curve against BoNTA HC (Figure 5.7) indicated that the specific binding is dependent on the concentration of [^{3}H] aminopterin. To determine the specific binding of each concentration of [^{3}H] aminopterin, the non-specific binding was deducted from total binding and then normalized between 0-100%. All samples were present in duplicate. The normalized data was used to generate the dose- response curve of [^{3}H] aminopterin against BoNTA HC. To determine the effective concentration of [^{3}H] aminopterin, a non-linear regression model (log (agonist) vs. normalized response-variable slope) was used. The EC_{50} of [^{3}H] aminopterin against BoNTA HC was estimated to be 703 ± 98 nM.

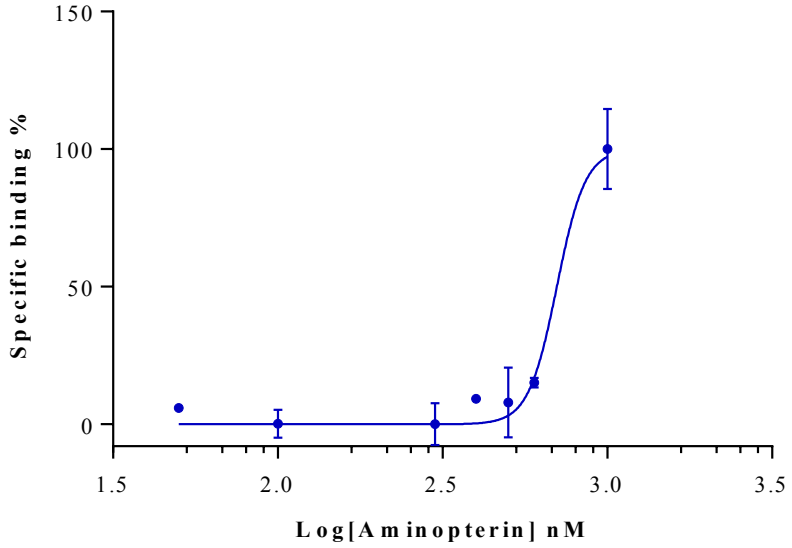


Figure 5.7 SPA binding curve for aminopterin against BoNTA HC. The EC_{50} of [^{3}H] aminopterin was determined to be 703 ± 98 nM using the SPA. The [^{3}H] aminopterin was used at variable concentrations in nM range and the concentration of SPA beads and BoNTA HC was constant for all the duplicate samples and controls at 2 mg/ml and 17 nM, respectively. Non-linear regression was used to plot the curve and the error bars correspond to standard deviations.

5.5.6 [^{3}H] Desmosine binding curve

We obtained EC_{50} for [^{3}H] desmosine against BoNTA HC from a SPA dose-response curve using variable concentration of [^{3}H] desmosine (Figure 5.8). The data in the graph was normalized between 0-100% after subtracting the background. The result demonstrated that increasing the concentration of [^{3}H] desmosine, the specific binding increases. The calculated EC_{50} for [^{3}H] desmosine against BoNTA HC is 1.6 ± 0.3 μM and. A non-linear regression model (log (agonist) vs. normalized response-Variable slope) was applied for estimating the EC_{50} of [^{3}H] desmosine against BoNTA HC.

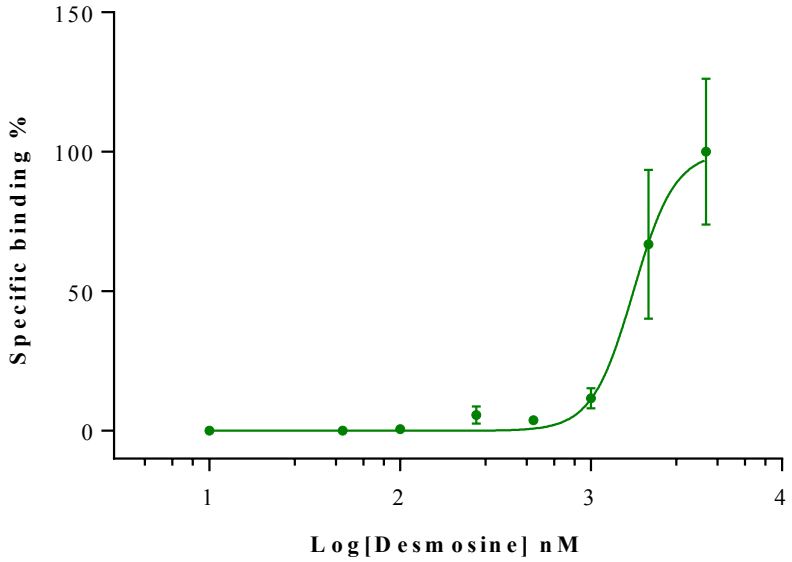


Figure 5.8 SPA dose-response curve for [^{3}H] desmosine against BoNTA HC. The EC_{50} value for [^{3}H] desmosine was estimated at 1.6 ± 0.3 µM using non-linear regression and model (log (agonist) vs. normalized response -variable slope). In this experiment, the concentration of SPA beads and BoNTA HC were kept constant at 2 mg/ml and 17 nM, respectively. The concentration of [^{3}H] desmosine varied from 4 µM to 10 nM. The specific binding was determined by subtracting the background from total binding. The error bars correspond to standard deviations.

5.5.7 Agreement between predicted and experimental binding profiles

The PSVLS-predicted binding profiles along with the SPA experimental results are summarized in Table 5.2. Aminopterin is predicted to bind preferentially to BoNTA HC, which is confirmed experimentally by SPA (EC_{50} (BoNTA HC) = 0.7 µM). Aminopterin did not score well in any region in the LC, consistent with SPA results that show no binding to BoNTA LC at 1µM and supported by previously reported inhibition assays at 10 µM[8]. Desmosine shows good holistic scores to both HC and LC. Binding to HC is confirmed experimentally by SPA, with an EC_{50} (BoNTA HC) of 1.6 µM. Desmosine was not found to bind to BoNTA LC in the SPA assay at 2 µM or lower. However, binding of desmosine to BoNTA LC is supported by the weak inhibition (25%) observed at 10 µM in our previous work[8]. According to holistic scores,

paclitaxel has a strong preference for BoNTA LC, which is confirmed by SPA with EC_{50} of 16.9 nM, and by a FRET inhibition assays (IC_{50} = 5.2 μM)[8]. The holistic scores for solanesyl-pyrophosphate suggest that this compound binds strongly to both BoNTA HC and LC, in agreement with SPA results at 0.6 μM. Fructose scores very poorly in all binding regions and has consistently produced a negative response in all our experiments. Solanesol did not pass the PC1 non-binder/non-binder threshold but its best region-specific holistic score in region 6, located in the LC, is above the region-specific threshold. No binding was observed for solanesol at 0.3 μM to either BoNTA HC or LC. We were unable to test solanesol at higher concentrations or in an inhibition assay to confirm its status. There is a very clear agreement between predicted and experimentally determined binding profiles for the selected ligands, indicating that PSVLS is an efficient methodology for the selective identification of non-orthosteric as well as orthosteric ligands.

Table 5.2 Predicted versus experimental binding profiles for selected ligands from PSVLS. Lowest positive concentration refers to the lowest concentration tested that gave a positive response. Highest concentration tested refers to the highest concentration tested that gave a negative response in ligand binding assay.

Compound Name	Computational Binding Preference (score)[1]	BoNTA LC	BoNTA HC	Lowest Positive Concentration	Highest Negative Concentration	EC_{50}	% Inhibition at 10μM (Dadgar et al., 2013)
Aminopterin	R27 (-4.7) HC TD R8 ***(-4.8)*** HC LBD R7 ***(-4.7)*** HC TC	−	+	1000 nM [4]	500 nM [3]	703 ± 98 nM	11
Desmosine	R6 (-5.5) LC CS R23 (-4.9) LC R10 (-4.8) HC TC R1 ***(-4.8)*** LC CS	−	+	2000 nM [4]	3000 nM [3]	1.6 ± 0.3μM	25
Solanesol	R6 (-6) LC	−	−	NA	300 nM	NA	NA
Sol.P.P	R27 (-13.8) HC TD R3 (-13.4) HC LBD R7 (-13.4) HC TC R2 (-12.9) LC R10 (-12.9) HC TC	+	+	600 nM [2]	NA	ND	NA
Paclitaxel	R14 (-11.0) LC near CS	+	−	50 nM	50 nM	16.91 nM	95
Fructose	R17***(-1.9)*** HC TD	−	−	NA	3000 nM	NA	0

[1] Score corresponds to holistic scores (Dadgar et al., 2013) at the best region(s) among the 33 regions scanned. Multiple regions are listed if the difference in binding energies was 1 unit or less. Scores not passing the per-region non-binder/binder threshold are shown in bold italic. [2] The lowest positive concentration of Sol.P.P was 600 nM for both BoNTA HC and LC
[3]Ligands were tested at higher concentration against BoNTA LC, producing a negative response; no lower concentrations were tested afterwards. [4] Positive ligands at higher concentration were tested against BoNTA HC at two lower concentrations (250 and 50 nM), producing negative responses.
NA = not applicable
ND = not determined
LC = light chain, including the catalytic site among others
HC = heavy chain
LBD = ligand binding domain in the HC
TD = translocation domain in the HC
CS = catalytic site in the LC

5.6 CONCLUSIONS

Target-specific molecular probes are organic compounds that bind to a particular protein (the target) and report its presence in the form of a detectable signal, such as fluorescence or scintillation caused by radioactivity. Molecular probes can be utilized in medical imaging to locate regions where the target protein is present or it is more/less abundant than normal, indicating a disease state. They can also be utilized in laboratory assays to identify and quantify a protein that is associated to a disease or condition. The current paradigm of molecular probes discovery is to modify existing medicinal drugs known to bind to the protein of interest. However, this approach has a number of disadvantages: not all medicinal drugs are suitable for labelling; labelling with an additional chemical group may negatively impact binding affinity for the target; the molecular probe may have biological activity against the target protein, precluding its use for monitoring the effects of novel drugs against the target; not all proteins for which a molecular probe is desirable have known ligands.

We applied PSVLS to identify molecular probes targeting non-orthosteric sites within the structure of BoNTA. The experimental binding profile obtained for 6 compounds tested independently against the light chain and the heavy chain of BoNTA is in excellent agreement with the binding profile predicted by PSVLS for these compounds.

Our results demonstrate that the PSVLS approach has the ability to identify probe candidates for multiple, spatially distinct binding regions within the structure of the target protein. These binding regions include orthosteric, allosteric, and other non-orthosteric sites. This enables the use of PSVLS for identification of probes that take advantage of unique non-conserved sites, thus increasing selectivity. Non-orthosteric binding sites are not under the same evolutionary pressure to be conserved as do orthosteric sites and, thus, present better

opportunities for the discovery and/or design of highly selective molecular probes. Moreover, some of the regions target by PSVLS may not interfere upon ligand binding with the main functional site of the target protein. This enables the use of the PSVLS probes for monitoring treatment response through medical imaging without the risk of affecting the actual treatment. Since PSVLS does not require prior knowledge of compounds that bind to the protein target of interest, we anticipate that this method will be able to identify molecular imaging probes for newly identified protein biomarkers from genomics/proteomics studies. The only condition for applying PSVLS to newly identified biomarkers is the availability of an experimental structure, or the possibility of modeling the biomarker's structure using well-established methods for protein structure modeling. The early identification of molecular probes will accelerate the validation of novel biomarkers by enabling scientists to develop target-specific assays to conduct their studies.

5.7 ACKNOWLEDGMENTS

This research was supported with funds from the Thunder Bay Regional Research Institute and the RBC Royal Bank's Dr. Mark Poznansky Mentorship Development Award. Initial virtual ligand screening was performed with support from the National Institutes of Health (NIH) (S06 GM053933). Additional computational utilized resources provided through SHARCNET (www.sharcnet.ca) and Lakehead University's High Performance Computing Centre (LUHPCC).

5.8 ABBREVIATIONS

BoNT, botulinum neurotoxin; BoNTA, botulinum neurotoxin type A; BoNTA LC, botulinum neurotoxin type A light chain; BoNTA HC, botulinum neurotoxin type A heavy chain;

PSVLS, protein scanning with virtual ligands screening; SPA, scintillation proximity assay; PVT, PolyVinyltoluene.

5.9 REFERENCES

1. (a) Lane, R.; Chubukov, P.; Liu, W.; Canals, M.; Cherezov, V.; Abagyan, R.; Stevens, R. C.; Katritch, V., Structure-Based Ligand Discovery Targeting Orthosteric and Allosteric Pockets of Dopamine Receptors. *Mol Pharmacol* **2013**; (b) Christopoulos, A., Allosteric binding sites on cell-surface receptors: novel targets for drug discovery. *Nat Rev Drug Discov* **2002,** *1* (3), 198-210; (c) Bond, C. J.; Jurica, M. S.; Mesecar, A.; Stoddard, B. L., Determinants of allosteric activation of yeast pyruvate kinase and identification of novel effectors using computational screening. *Biochemistry* **2000,** *39* (50), 15333-43; (d) McKay, D. B.; Chang, C.; Gonzalez-Cestari, T. F.; McKay, S. B.; El-Hajj, R. A.; Bryant, D. L.; Zhu, M. X.; Swaan, P. W.; Arason, K. M.; Pulipaka, A. B.; Orac, C. M.; Bergmeier, S. C., Analogs of methyllycaconitine as novel noncompetitive inhibitors of nicotinic receptors: pharmacological characterization, computational modeling, and pharmacophore development. *Mol Pharmacol* **2007,** *71* (5), 1288-97; (e) Ramirez, U. D.; Myachina, F.; Stith, L.; Jaffe, E. K., Docking to large allosteric binding sites on protein surfaces. *Adv Exp Med Biol* **2010,** *680*, 481-8; (f) Arencibia, J. M.; Pastor-Flores, D.; Bauer, A. F.; Schulze, J. O.; Biondi, R. M., AGC protein kinases: from structural mechanism of regulation to allosteric drug development for the treatment of human diseases. *Biochim Biophys Acta* **2013,** *1834* (7), 1302-21; (g) Liu, F.; Li, F.; Ma, A.; Dobrovetsky, E.; Dong, A.; Gao, C.; Korboukh, I.; Liu, J.; Smil, D.; Brown, P. J.; Frye, S. V.; Arrowsmith, C. H.; Schapira, M.; Vedadi, M.; Jin, J., Exploiting an allosteric binding site of PRMT3 yields potent and selective inhibitors. *J Med Chem* **2013,** *56* (5), 2110-24; (h) Tomita, N.; Hayashi, Y.; Suzuki, S.; Oomori, Y.; Aramaki, Y.; Matsushita, Y.; Iwatani, M.; Iwata, H.; Okabe, A.; Awazu, Y.; Isono, O.; Skene, R. J.; Hosfield, D. J.; Miki, H.; Kawamoto, T.; Hori, A.; Baba, A., Structure-based discovery of cellular-active allosteric inhibitors of FAK. *Bioorg Med Chem Lett* **2013,** *23* (6),

1779-85; (i) Mahmoud, M. M.; Ali, H. I.; Ahn, K. H.; Damaraju, A.; Samala, S.; Pulipati, V. K.; Kolluru, S.; Kendall, D. A.; Lu, D., Structure-Activity Relationship Study of Indole-2-carboxamides Identifies a Potent Allosteric Modulator for the Cannabinoid Receptor 1 (CB1). *J Med Chem* **2013**; (j) Shiryaev, S. A.; Cheltsov, A. V.; Gawlik, K.; Ratnikov, B. I.; Strongin, A. Y., Virtual ligand screening of the National Cancer Institute (NCI) compound library leads to the allosteric inhibitory scaffolds of the West Nile Virus NS3 proteinase. *Assay Drug Dev Technol* **2011,** *9* (1), 69-78; (k) Laine, E.; Martinez, L.; Ladant, D.; Malliavin, T.; Blondel, A., Molecular motions as a drug target: mechanistic simulations of anthrax toxin edema factor function led to the discovery of novel allosteric inhibitors. *Toxins (Basel)* **2012,** *4* (8), 580-604.

2. Lebeda, F. J.; Cer, R. Z.; Mudunuri, U.; Stephens, R.; Singh, B. R.; Adler, M., The Zinc-Dependent Protease Activity of the Botulinum Neurotoxins. *Toxins* **2010,** *2* (5), 978-997.

3. Fasshauer, D.; Sutton, R. B.; Brunger, A. T.; Jahn, R., Conserved structural features of the synaptic fusion complex: SNARE proteins reclassified as Q- and R-SNAREs. *P Natl Acad Sci USA* **1998,** *95* (26), 15781-15786.

4. Sharma, S. K.; Ferreira, J. L.; Eblen, B. S.; Whiting, R. C., Detection of type A, B, E, and F Clostridium botulinum neurotoxins in foods by using an amplified enzyme-linked immunosorbent assay with digoxigenin-labeled antibodies. *Appl Environ Microb* **2006,** *72* (2), 1231-1238.

5. Ahmed, S. A.; Olson, M. A.; Ludivico, M. L.; Gilsdorf, J.; Smith, L. A., Identification of residues surrounding the active site of type A botulinum neurotoxin important for substrate recognition and catalytic activity. *The protein journal* **2008,** *27* (3), 151-62.

6. Dhaked, R. K.; Singh, M. K.; Singh, P.; Gupta, P., Botulinum toxin: Bioweapon & magic drug. *Indian J Med Res* **2010,** *132* (5), 489-503.

7. (a) Eubanks, L. M.; Dickerson, T. J.; Janda, K. D., Technological advancements for the detection of and protection against biological and chemical warfare agents. *Chem Soc Rev* **2007,** *36* (3), 458-470; (b) Ward, A. B.; Molenaers, G.; Colosimo, C.; Berardelli, A., Clinical value of botulinum toxin in neurological indications. *Eur J Neurol* **2006,** *13*, 20-26.

8. Dadgar, S.; Ramjan, Z.; Floriano, W. B., Paclitaxel Is an Inhibitor and Its Boron Dipyrrornethene Derivative Is a Fluorescent Recognition Agent for Botulinum Neurotoxin Subtype A. *J Med Chem* **2013,** *56* (7), 2791-2803.

9. (a) Floriano, W. B.; Vaidehi, N.; Zamanakos, G.; Goddard, W. A., 3rd, HierVLS hierarchical docking protocol for virtual ligand screening of large-molecule databases. *J Med Chem* **2004,** *47* (1), 56-71; (b) Avila, A. Eshu: A Set of Analysis Tools For The Orunmila Ligand-Protein Docking And Scoring System. California State Polytechnic University Pomona, Pomona, 2009.

10. Weatherly, G. T.; Bouvier, A.; Lydiard, D. D.; Chapline, J.; Henderson, I.; Schrimsher, J. L.; Shepard, S. R., Initial purification of recombinant botulinum neurotoxin fragments for pharmaceutical production using hydrophobic charge induction chromatography. *Journal of chromatography. A* **2002,** *952* (1-2), 99-110.

11. Satta, J.; Laurila, A.; Paakko, P.; Haukipuro, K.; Sormunen, R.; Parkkila, S.; Juvonen, T., Chronic inflammation and elastin degradation in abdominal aortic aneurysm disease: an immunohistochemical and electron microscopic study. *European journal of vascular and endovascular surgery : the official journal of the European Society for Vascular Surgery* **1998,** *15* (4), 313-9.

12. Seddon, J.; Kasprowicz, V.; Walker, N. F.; Yuen, H. M.; Sunpath, H.; Tezera, L.; Meintjes, G.; Wilkinson, R. J.; Bishai, W. R.; Friedland, J. S.; Elkington, P. T., Procollagen III

N-terminal Propeptide and Desmosine are Released by Matrix Destruction in Pulmonary Tuberculosis. *The Journal of infectious diseases* **2013,** *208* (10), 1571-9.

13. (a) Turino, G. M.; Ma, S.; Lin, Y. Y.; Cantor, J. O.; Luisetti, M., Matrix elastin: a promising biomarker for chronic obstructive pulmonary disease. *American journal of respiratory and critical care medicine* **2011,** *184* (6), 637-41; (b) Maclay, J. D.; McAllister, D. A.; Rabinovich, R.; Haq, I.; Maxwell, S.; Hartland, S.; Connell, M.; Murchison, J. T.; van Beek, E. J.; Gray, R. D.; Mills, N. L.; Macnee, W., Systemic elastin degradation in chronic obstructive pulmonary disease. *Thorax* **2012,** *67* (7), 606-12.

14. (a) Devenport, N. A.; Reynolds, J. C.; Parkash, V.; Cook, J.; Weston, D. J.; Creaser, C. S., Determination of free desmosine and isodesmosine as urinary biomarkers of lung disorder using ultra performance liquid chromatography-ion mobility-mass spectrometry. *Journal of chromatography. B, Analytical technologies in the biomedical and life sciences* **2011,** *879* (32), 3797-801; (b) Lamerz, J.; Friedlein, A.; Soder, N.; Cutler, P.; Dobeli, H., Determination of free desmosine in human plasma and its application in two experimental medicine studies. *Anal Biochem* **2013,** *436* (2), 127-36.

15. Smith, A.; Hum, M.; Winick, N. J.; Kamen, B. A., A case for the use of aminopterin in treatment of patients with leukemia based on metabolic studies of blasts in vitro. *Clin Cancer Res* **1996,** *2* (1), 69-73.

6 CHAPTER SIX: GENERAL DISCUSSION

Botulism is a severe neurological disease cause by the botulinum neurotoxin. The BoNTA is the most potent one among the 7 serotypes and it is the cause of botulism in humans.

Paclitaxel identified as an inhibitor (IC_{50} equal to 5.2 μM) for the zinc protease activity of BoNTA LC using FRET assay. Also the specific binding of [^{3}H] paclitaxel to BoNTA LC was verified using SPA assay. The fluorescent derivative of paclitaxel (PAC-BDP) exhibited binding interaction to both BoNTA LC and BoNTA HC using FP. However PAC-BDP did not compete with PAC for the active site. The $EC_{50}$of [^{3}H] DES (1.6 ± 0.3 μM) and [^{3}H] AMN (703 ± 98 nM) against BoNTA HC were obtained using SPA. Additionally, we confirmed the binding of DES and AMN to BoNTA HC using FP competition assay. [^{3}H] Solanesyl PP is the other radiolabeled compound which showed strong binding affinity to both BoNTA LC and HC using SPA assay.

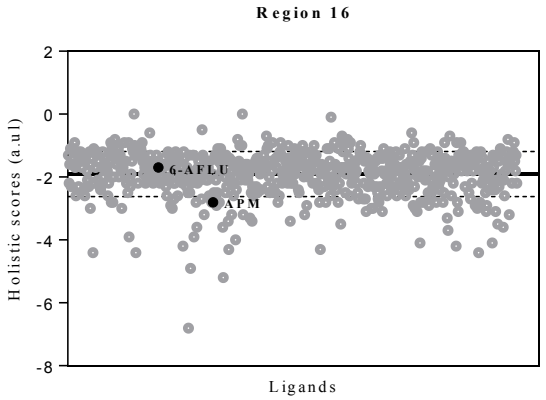
Protein sequence and structural analysis of different serotypes of BoNT (A-G) showed similarity and identity percentages of more than 50% and 31%, respectively. This analysis allowed the identification of critical binding pockets that are important to detect BoNTA from the other serotypes. The critical binding pockets R16, R23, R8, R30, R13, R3, R19, R21, and R22 were explored computationally to identify two novel fluorescent recognition agents, APM-BDP and 6-AFLU for BoNTA detection. The binding of APM-BDP to BoNTA LC and 6-AFLU to BoNTA HC was confirmed experimentally using FP. The EC_{50} for APM-BDP and 6-AFLU were estimated from dose-response curves to be equal to 20.96 ± 10 nM and 546 ± 60nM, respectively. The binding of APM and APM-BDP at the same site in BoNTA LC was verified using a FP competition assay.

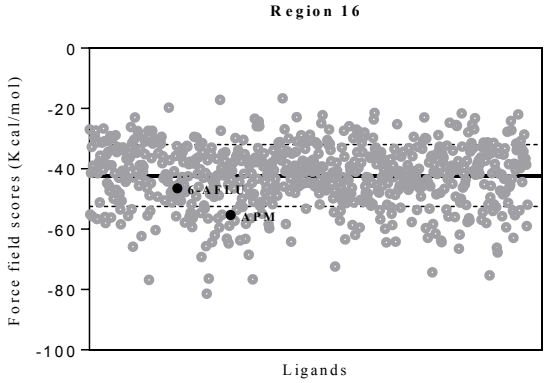
The development of an easy, rapid and reliable technique for detection of BoNT in contaminated food and body samples is very desirable since MLA is the standard procedure to detect BoNTs. BoNTA detection in the primary stage of contamination is critical since the antitoxin (the only treatment available) can only be effective if given before entrance of the toxin into the nerve cells. The detection time of MLA is between 2-4 days which is the major limitation of this technique. MLA required animal facility which makes it more expensive and complicated compared to FP. Simplicity of FP assay makes it suitable for high throughput screening of large number of samples. FP technique is also suitable for portable devices. In this research, the strength of FP assay for detection was confirmed by testing the three identified fluorescent probes against BoNTA LC and HC individually, and testing APM-BDP and PAC-BDP against BoNTA LC and HC in competition assays. In addition, most studies reported in the literature are focused on detection of BoNTA based on its catalytic activity. However the active site is only accessible after activation of the toxin (cleavage of disulfide bond).Therefore the detection of BoNTA using recognition agents that bind to other binding pockets of BoNTA is advantageous. The identified fluorescent probes APM-BDP, 6-AFLU and PAC-BDP can be used in a multi-probe FP assay to detect BoNTA complex since they bind to the different binding sites within the BoNT structure. The application of paclitaxel derivatives in microtubule imaging is a limitation of using Paclitaxel-BDP as a recognition agent for BoNTA in cell-based assays. However combination of all three fluorescent probes in a multi-probe assay can reduce the incidence of false positives due to probe interactions with other proteins.

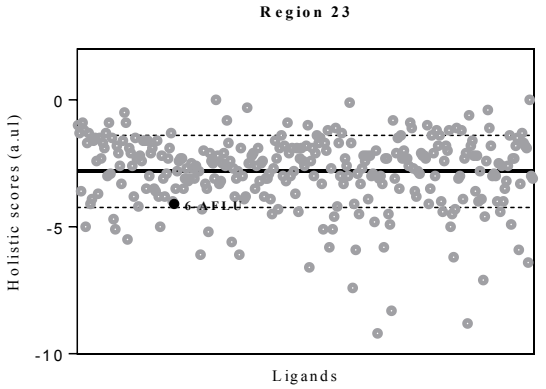
Appendices

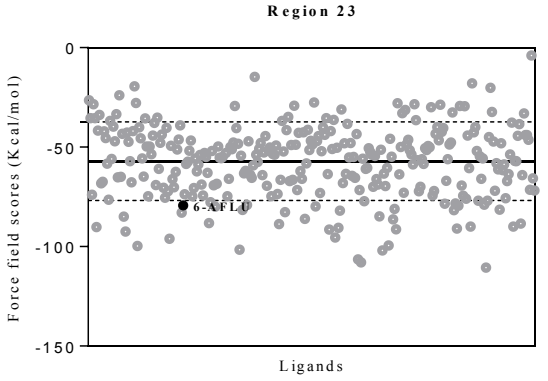
Appendix A

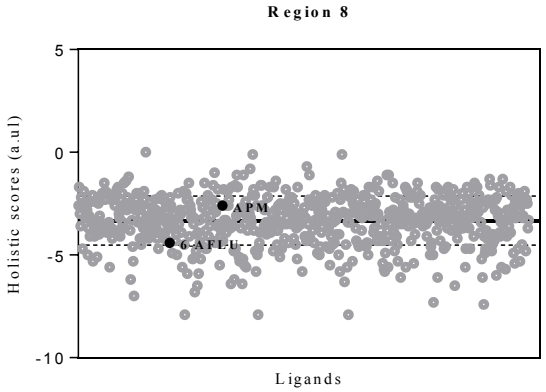
(A-1) The holistic and force field binding scores of unlabeled ligands at the critical binding regions.

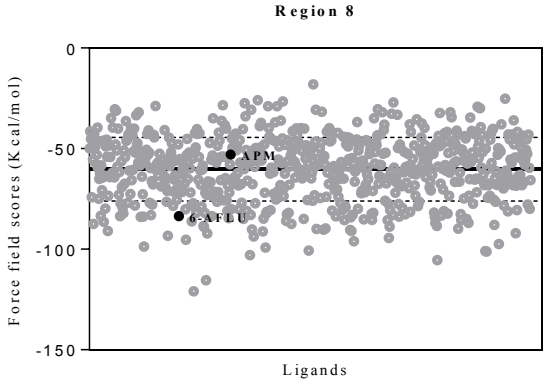




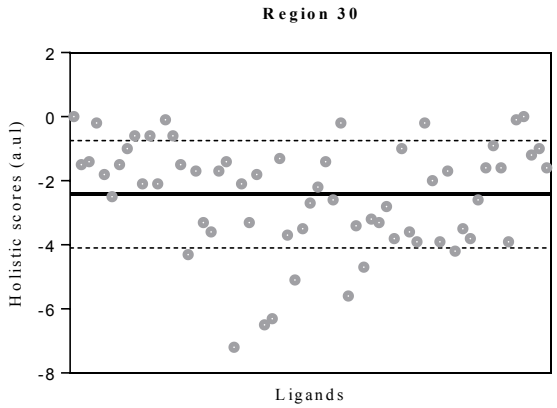


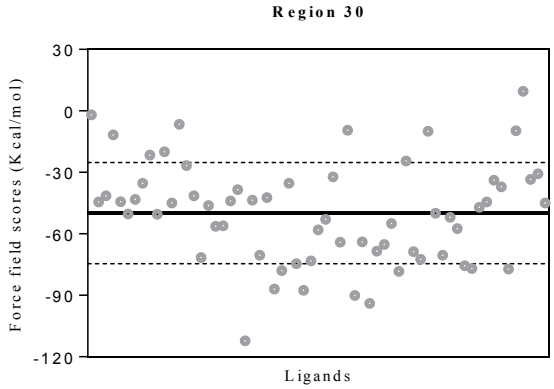


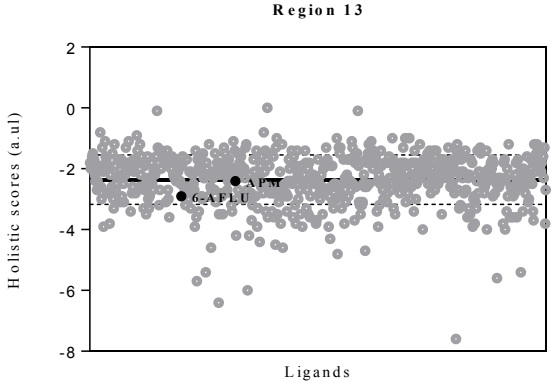


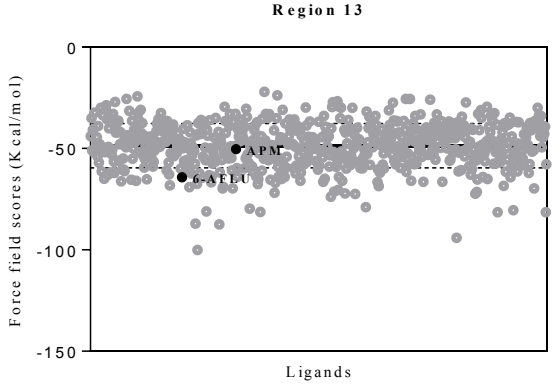


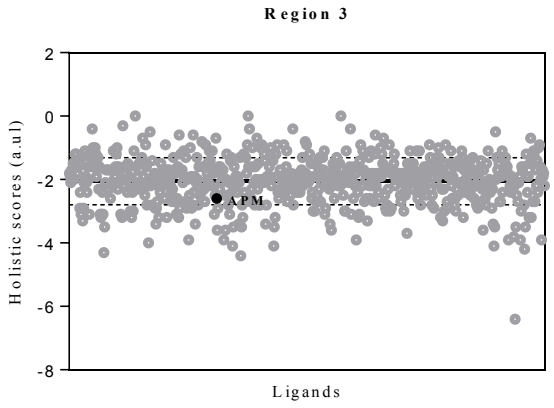
(A-2) The holistic and force field binding scores of unlabeled ligands at the critical binding regions.

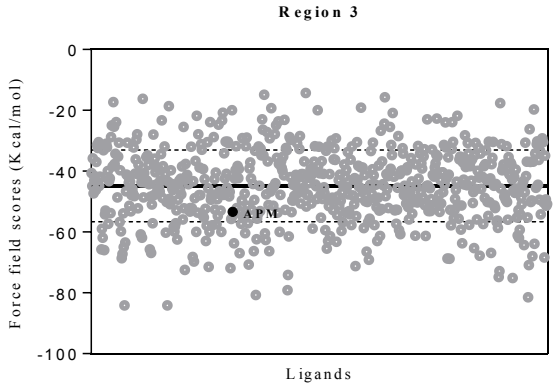




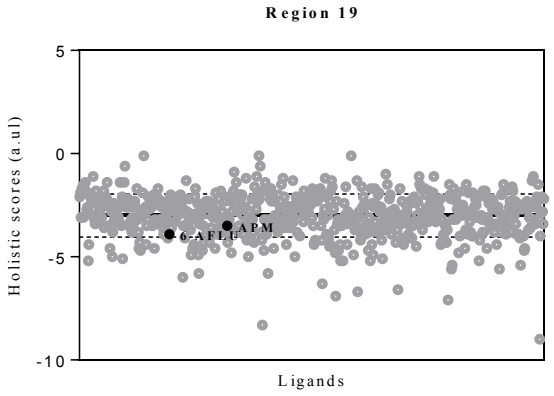


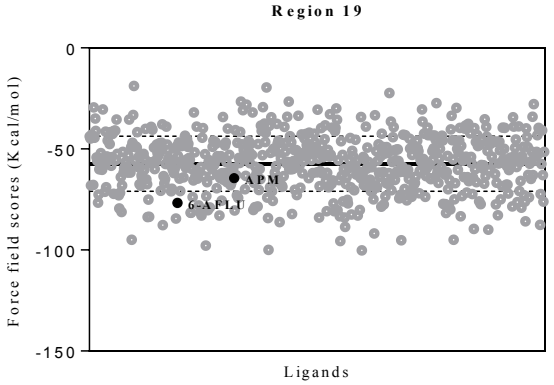


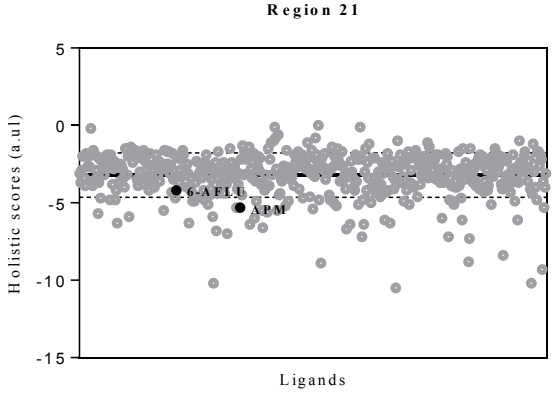


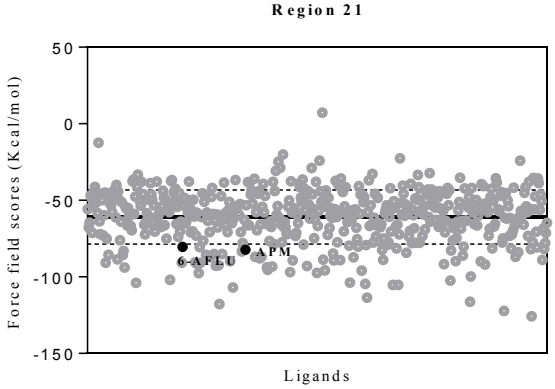


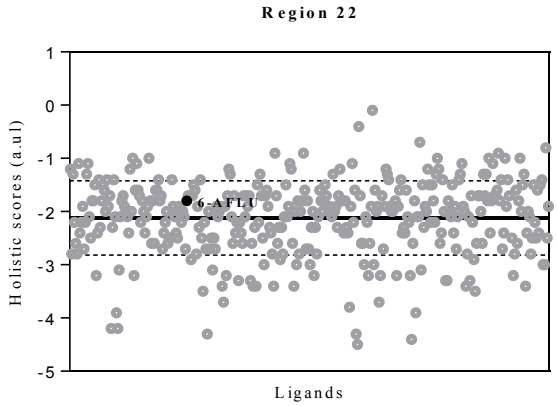
(A-3) The holistic and force field binding scores of unlabeled ligands at the critical binding regions.

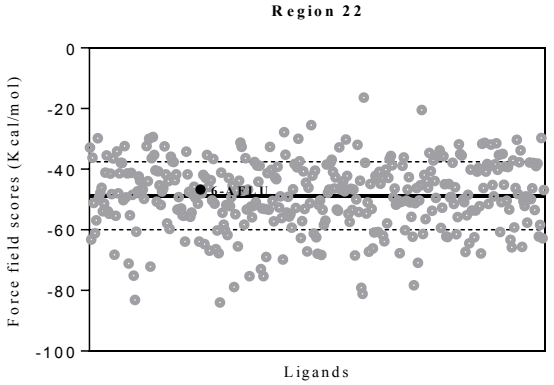






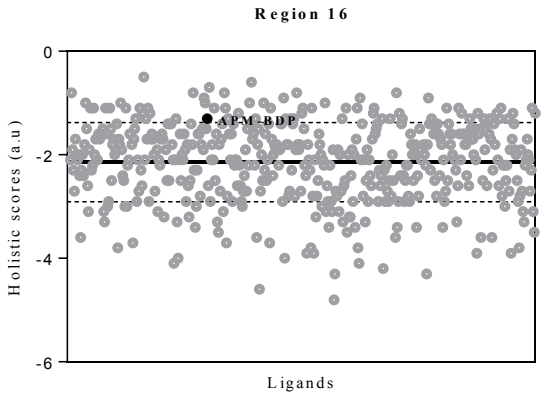


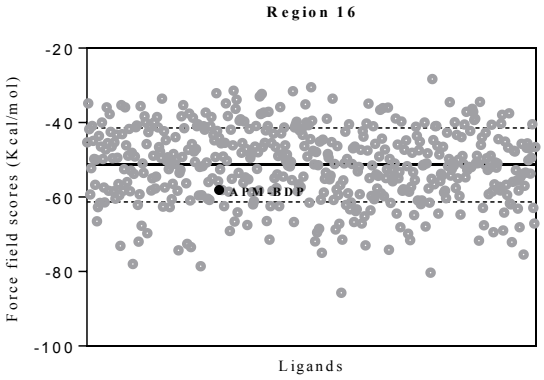


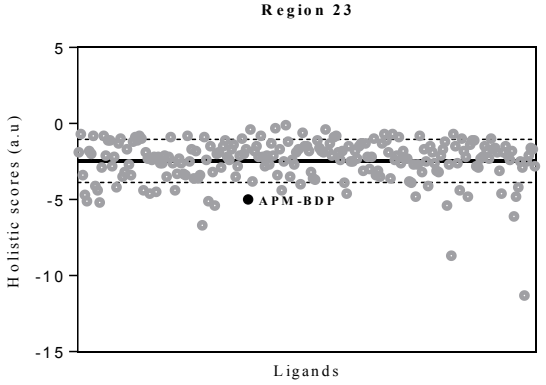


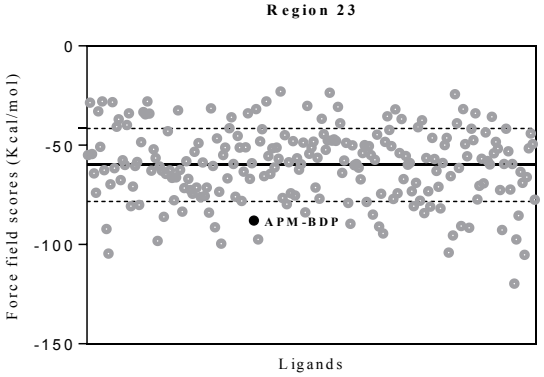
Appendix B

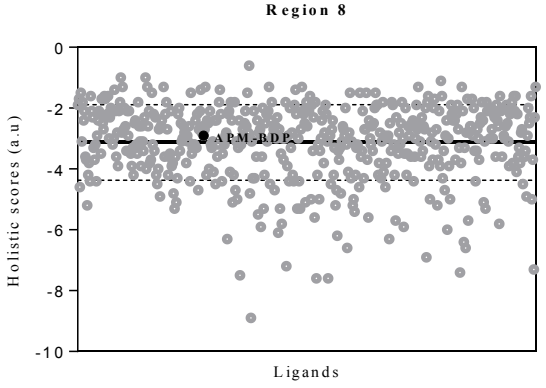
(B-1) The holistic and force field binding scores of fluorescent labeled ligands at the critical binding regions.

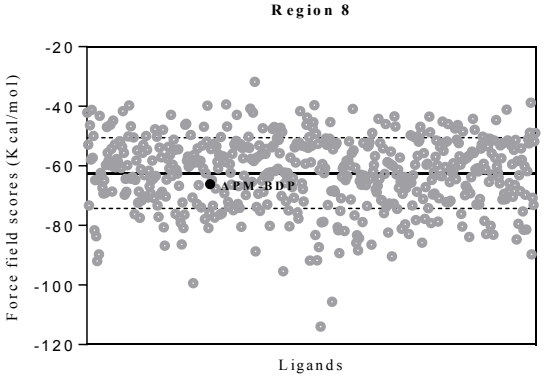




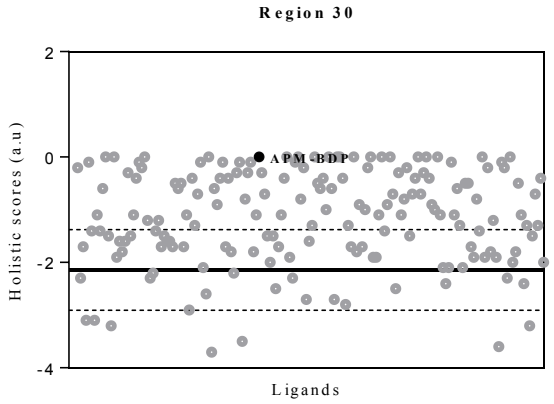


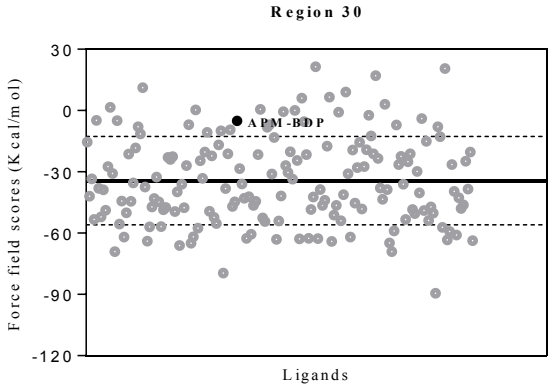


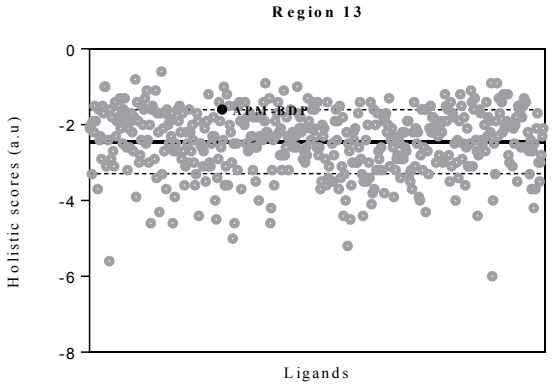


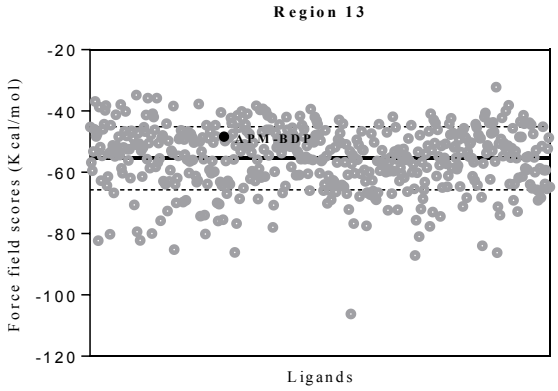


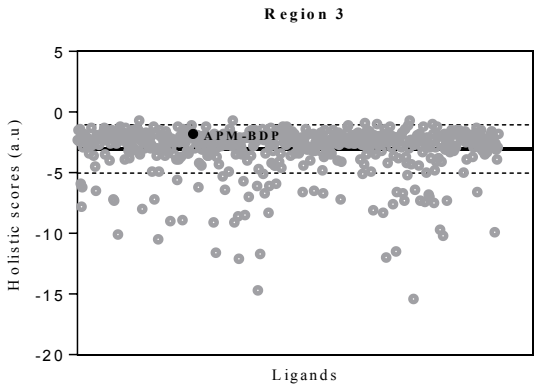
(B-2) The holistic and force field binding scores of fluorescent labeled ligands at the critical binding regions.

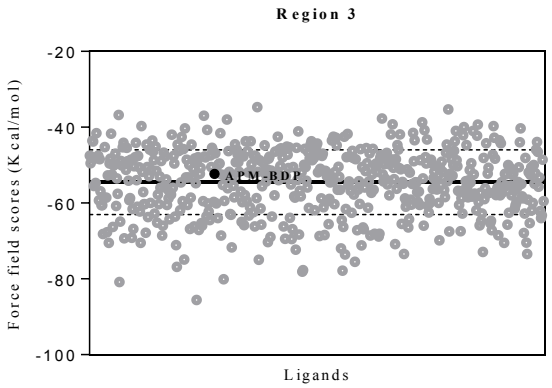




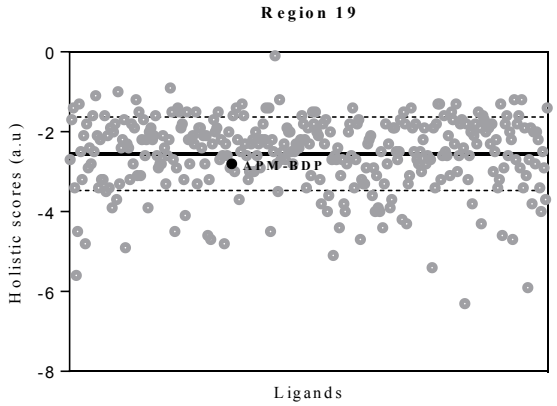


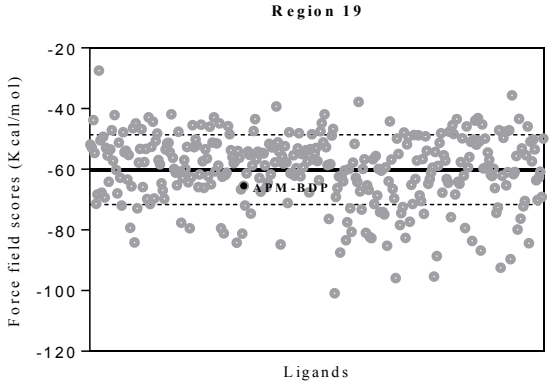


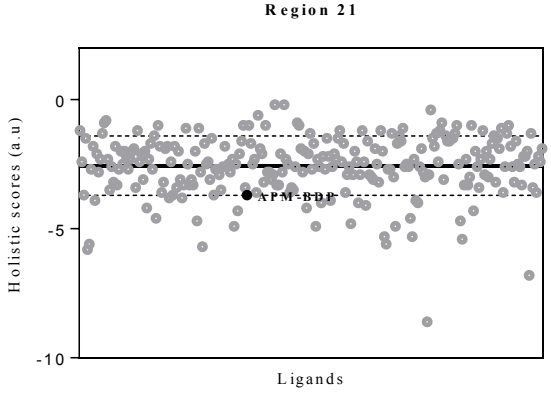


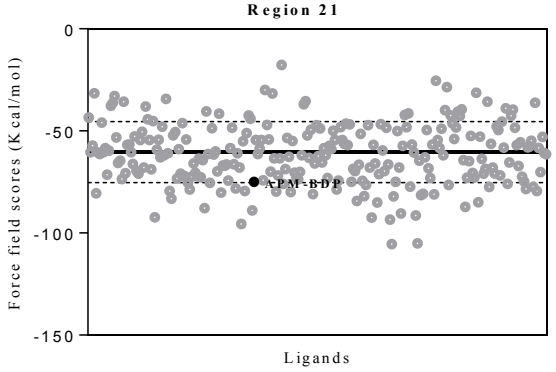


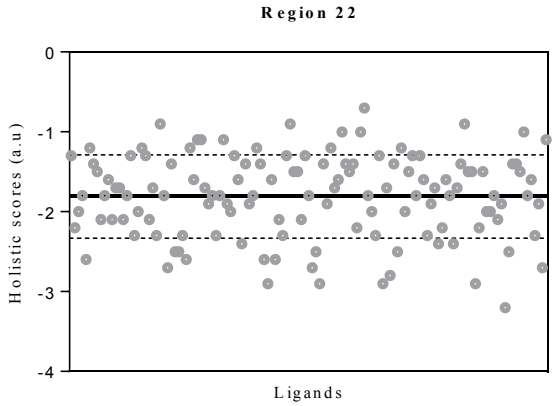
(B-3) The holistic and force field binding scores of fluorescent labeled ligands at the critical binding regions.

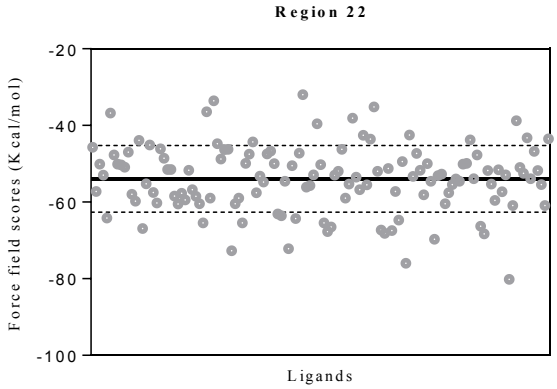






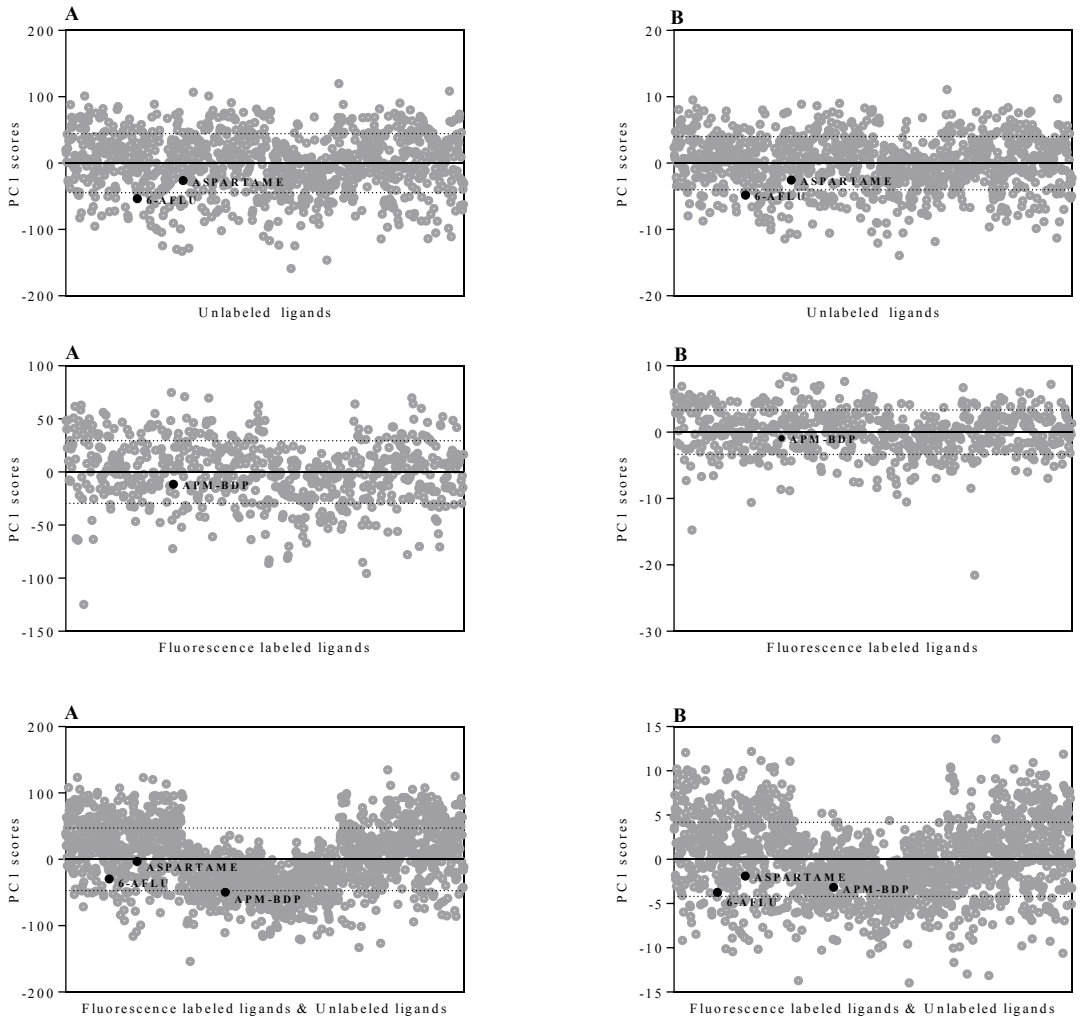






Appendix C

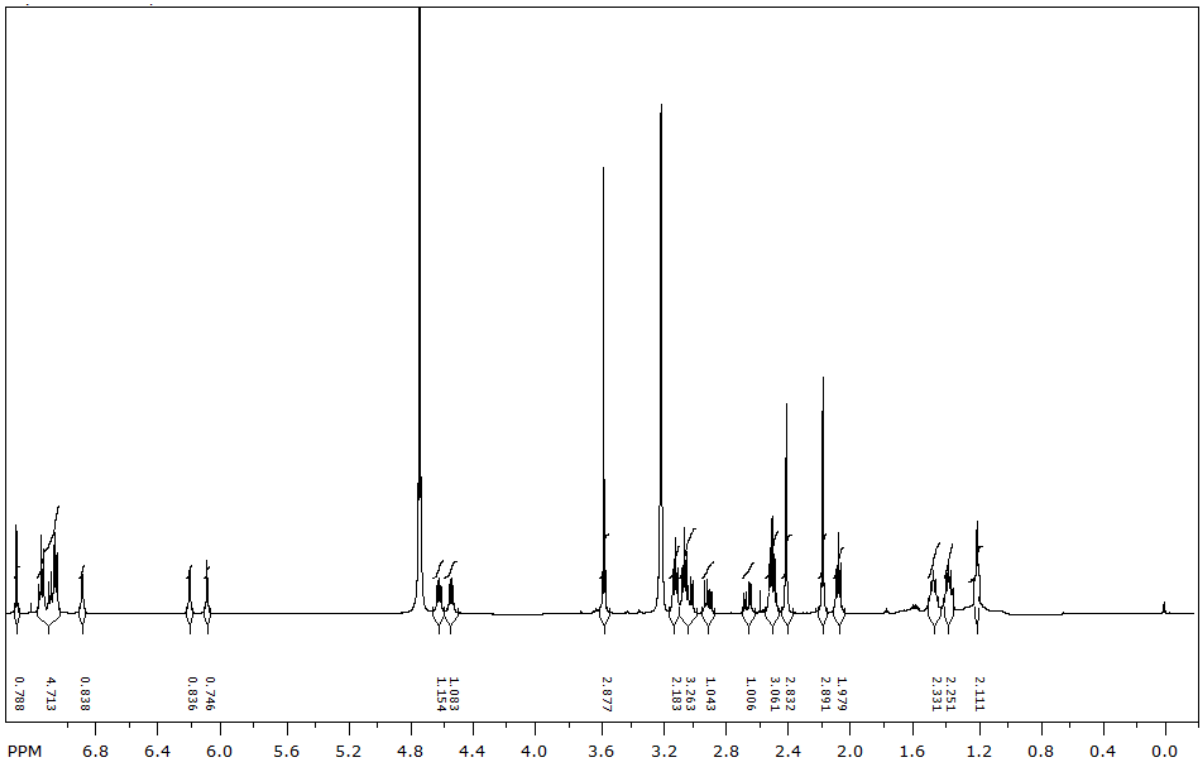
PCA scatter plots of the molecular docking of unlabeled and fluorescence labeled libraries against BoNTA



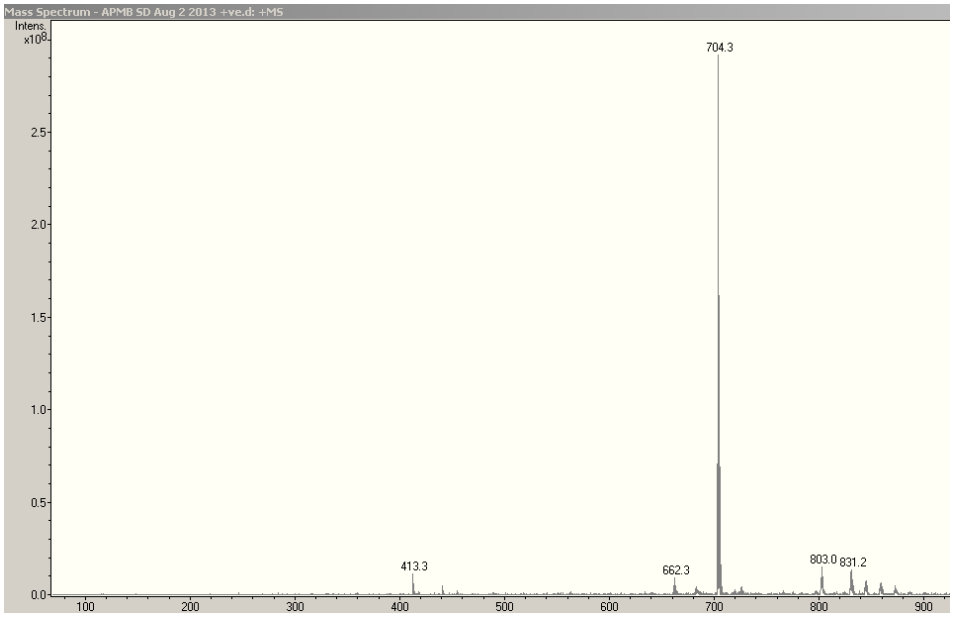
Screened library	% variance for PC1 Matrix: Var-covar Missing values: Mean value imputation (A)	% variance for PC1 Matrix: correlation Missing values: iterative imputation (B)
Unlabeled library	46.991	48.705
Fluorescence labeled library	26.997	33.76
Unlabeled and fluorescence labeled libraries	47.064	53.144

Appendix D

NMR spectrum of APM-BDP

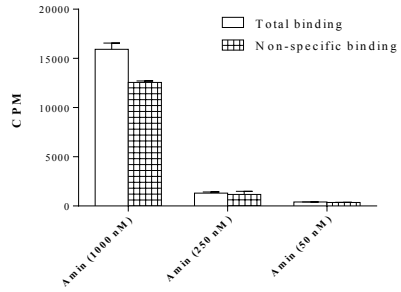
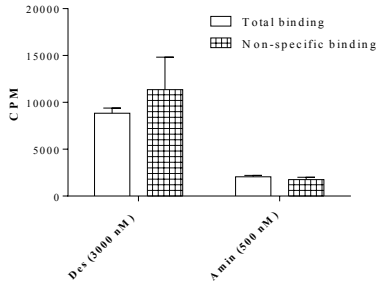
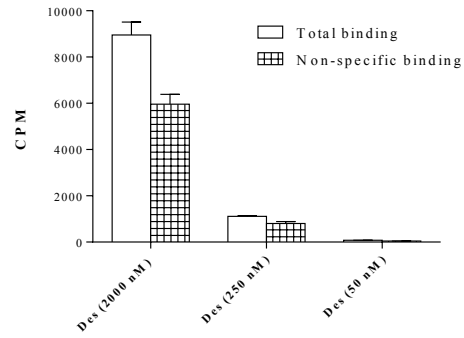
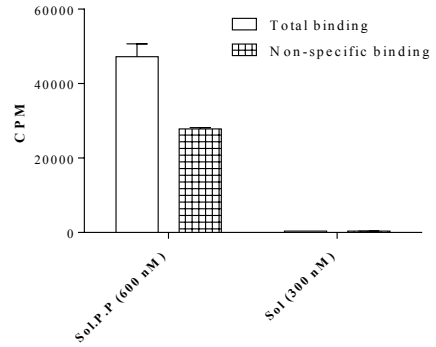
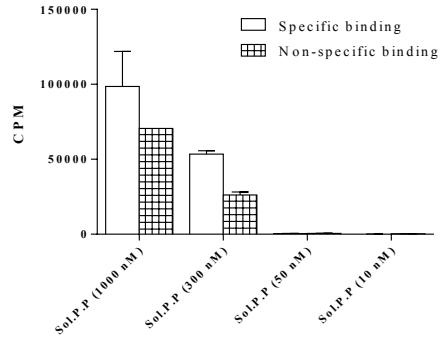
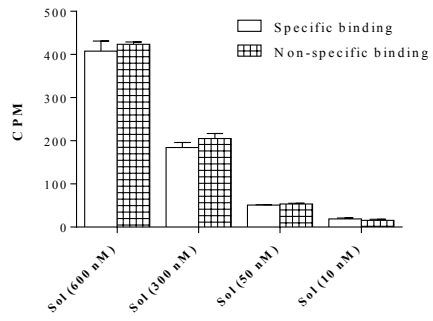


LC/MS spectrum of APM-BDP



Appendix E

Total binding versus non-specific binding of radio-ligands against BoNTA LC and HC

BoNTA HC	BoNTA LC
	
	
* 	* 
	* 

*These graphs were generated using data before correction of Chromalink effect.